U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

PhD

3.º
CICLO

FCUP
UA
UMinho
2016

U.PORTO

Non-Gaussian Models for the Formation of the Large Scale Structure in the Universe

Arlindo Miguel Múrias Trindade

FC

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

universidade de aveiro

Universidade do Minho

Non-Gaussian Models for the Formation of the Large Scale Structure in the Universe

Arlindo Miguel Múrias Trindade

Tese de Doutoramento apresentada à
Faculdade de Ciências da Universidade do Porto
Física
2016

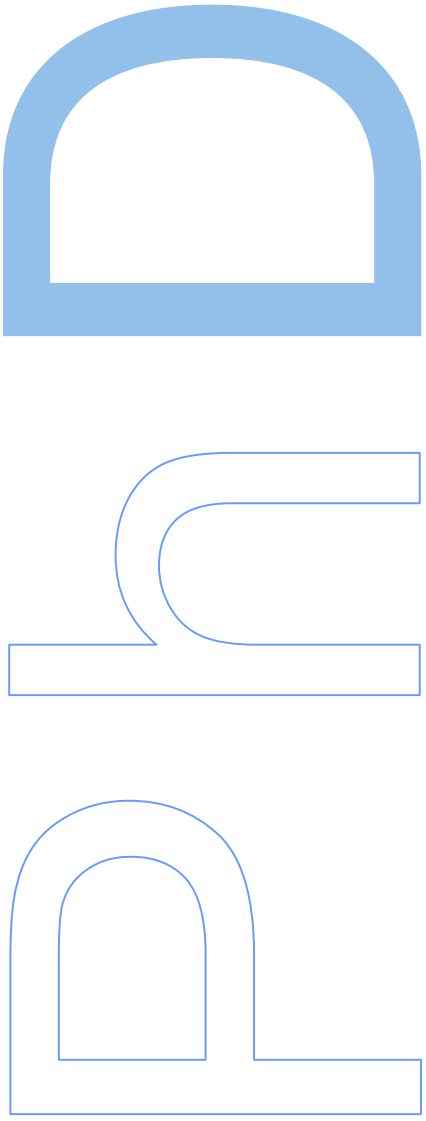
PhD

Non-Gaussian Models for the Formation of the Large Scale Structure in the Universe

Arlindo Miguel Múrias Trindade
Doutoramento em Física
Departamento de Física e Astronomía
Faculdade de Ciências da Universidade do Porto (FCUP)
2016

Orientador
Pedro T. P. Viana, Professor Auxiliar, FCUP

Coorientador
Pedro P. Avelino, Professor Auxiliar, FCUP



Non-Gaussian Models for the Formation of the Large Scale Structure in the Universe

By

Arlindo Miguel Múrias Trindade

Departamento de Física e Astronomía

Faculdade de Ciências da Universidade do Porto

Tese submetida à Faculdade de Ciências da Universidade do Porto

para obtenção do grau de Doutor em Física

Orientadores: P. T. P. Viana & P. P. Avelino

2016

Abstract

The redshift dependence of the abundance of galaxy clusters is very sensitive to the statistical properties of the primordial density perturbations. Therefore, it can be used to probe small deviations from Gaussian initial conditions. Such deviations constitute a very important signature of many inflationary scenarios, and are thus expected to provide crucial information on the physical processes which took place in the very early Universe.

In this thesis, we have studied how primordial non-Gaussianities may affect the estimation of cosmological parameters. We have found that, when the information contained in galaxy cluster counts is used to reconstruct the dark energy equation of state parameter as a function of redshift, assuming erroneously that no primordial non-Gaussianities exist, an apparent evolution with time in the effective dark energy equation of state, characterized by the appearance of a clear discontinuity, may arise. Furthermore, we have also determined the magnitude of the biases which may be introduced in the estimation of a broader set of cosmological parameters by wrongly assuming the absence of primordial non-Gaussianities. Using cluster counts, we find that the estimation of the present-day dark energy density and its equation of state parameter are not very sensitive to the non-Gaussian properties of the density field. On the other hand, we show that the biases can be considerably larger in the estimation of the amplitude of the primordial density perturbations. The results suggest that a significant level of non-Gaussianity at cluster scales may help reconciling the constraints on the amplitude of the primordial perturbations obtained using galaxy cluster number counts from the Planck Sunyaev-Zel'dovich Catalog with that obtained from the primary Cosmic Microwave Background anisotropies measured by the Planck satellite.

Galaxy clusters are a valuable source of cosmological information, since their formation and evolution strongly depends on the underlying cosmology and on the statistical nature of the primordial density fluctuations. However, before they can be used as cosmological probes, the non-trivial task of accurately estimating their mass must be done. The internal properties of galaxy cluster are usually used as mass proxies to accomplish such task. In this thesis, we also investigated, for the first time, the impact of primordial non-Gaussianities (PNG) on the scaling properties of galaxy clusters. We performed a series of cosmological hydrodynamic N-body simulations featuring adiabatic gas physics and different levels of non-Gaussian initial conditions within the ΛCDM framework. We focus on the $T - M$, $S - M$, $Y - M$ and $Y_X - M$ scalings relating the total cluster mass with temperature, entropy and SZ cluster integrated pressure that reflect the thermodynamical state of the intra-cluster medium. Our results show that PNG have an impact on cluster scalings laws. The mass power-law indexes of the scalings are almost unaffected by the existence of PNG but the amplitude and redshift evolution of their normalizations are clearly affected. The effect is stronger for the evolution of the $Y - M$ and $Y_X - M$ normalizations, which change by as much as 20%

and 13% when f_{NL} (parameter that controls the level of non-Gaussianity) varies from -500 to 500, respectively. These results are consistent with the view that positive/negative f_{NL} affects cluster profiles due to an increase/decrease of cluster concentrations. At low values of f_{NL}, as suggested by present Planck constraints on a scale invariant f_{NL}, the impact on the scalings normalizations is only a few percent, which is small when compared with the effect of additional gas physics and other cosmological effects such as dark energy. However, if f_{NL} is in fact a scale dependent parameter, PNG may have larger positive/negative amplitudes at clusters scales and therefore our results suggest that PNG should be taken into account when galaxy cluster data is used to infer cosmological parameters or to assess the constraining power of future cluster surveys.

Finally, we present the COSMOABC package, a Python Approximate Bayesian Computation (ABC) sampler featuring a *Population Monte Carlo* variation of the original ABC algorithm. The code, which uses an adaptive importance sampling scheme, is very flexible, highly parallelized and can be easily coupled to an external simulator, while allowing to incorporate arbitrary distance and prior functions. As an example of a practical application, we have coupled COSMOABC with the NUMCOSMO library and demonstrate how it can be used to estimate posterior probability distributions of cosmological parameters based on measurements of galaxy cluster number counts without computing the likelihood function.

Resumo

A abundância de enxames de galáxias como função do desvio para o vermelho depende fortemente das propriedades estatísticas da perturbações de densidade primordiais e, portanto, podem ser usadas para medir pequenos desvios da Gaussianidade nas condições iniciais. Tais desvios, constituem uma assinatura importante de inúmeros modelos inflacionários, pelo que é espectável que forneçam informações cruciais sobres os processos físicos que tiveram lugar no Universo primordial.

Nesta tese, estudámos a forma como a existência de não-Gaussianidades primordiais pode afectar a estimação de parâmetros cosmológicos. O uso da informação contida nas contagens de enxames de galáxias para reconstruir a equação de estado de energia escura como função do desvio para o vermelho, assumindo erradamente a inexistência de não-Gaussianidades primordiais, pode conduzir a uma evolução aparente da equação de estado da energia escura, caracterizada por uma clara descontinuidade. Calculámos também a magnitude dos desvios introduzidos na estimação de um conjunto mais alargado de parâmetros cosmológicos, se assumirmos erradamente a inexistência de não-Gaussianidades primordiais. Verificámos que, embora a estimação da densidade de energia escura hoje e respectiva equação de estado não sejam muito sensíveis às propriedades não-Gaussianas do campo de densidades primordial, constatou-se que a magnitude dos desvios é consideravelmente maiore no caso da amplitude das fluctuações primordiais. Estes resultados sugerem que um nível significativo de não-Gaussianidade à escala dos enxames de galáxias permitiriam reconciliar os constrangimentos observacionais na amplitude das fluctuações de densidade primordial obtidos usando contagens de enxames de galáxias do catálogo Sunyaev-Zel'dovich do Planck com os constrangimentos obtidos recorrendo a medições das anisotropias primárias da radiação cósmica de microondas efectuadas pelo Satélite Planck. Os enxames de galáxis são uma valiosa fonte de informação cosmológica, dado que a sua formação e evolução dependem fortemente do modelo cosmológico e na natureza estatística das perturbações de densidade primordiais. Contudo, antes de poderem ser usados como sondas cosmológicas, a tarefa não-trivial de estimar como precisão a massa destes objectos tem de ser feita. As propriedades internas dos enxames de galáxias são geralmente usadas para concretizar tal tarefa. Nesta tese, investigámos, pela primeira vez, o impacto das não-Gaussianidades primordiais nas relações de escala dos enxames de galáxias. Realizámos uma serie de simulações hidrodinâmicas de N-corpos com fisica de gás adiabático e com diferentes nívels de condições iniciais não-Gaussianas, no contexto do modelo ΛCDM. Focámo-nos nas relações de escala $T-M$, $S-M$, $Y-M$ and Y_X-M, que relacionam a massa total do enxame com a temperatura, a entropia, a pressão integrada SZ do enxame que reflecte o estado termodinâmico do meio intra-enxame. Os resultados mostam que as não-Gaussianidades primordiais têm de facto impacto

nas relações de escala dos enxames. Embora os indices *power-law* das relações de escala não sejam afectados pela existencia de não-Gaussianidades, a amplitude e evolução com o desvio para o vermelho das correspondentes normalizações são. O efeito é mais forte na evolução das normalizações de $Y - M$ and $Y_X - M$, que podem variar cerca de 22% e 16% respectivamente, quando f_{NL} (o parametro que controla o nível de não-Gaussianidade) varia de -500 a 500. Estes resultados estão de acordo com o facto que valores positivos/negativos de f_{NL} afectam os prefis dos enxames resultande do aumento/diminuição das concentrações. Para valores baixos de f_{NL}, sugeridos pelo constrangimentos obtivos pela colaboração Planck para um f_{NL} constante com a escala, o impacto nas normalizações das relações de escala é de poucos por cento, pelo que é pequeno quando comparado com outra fisica de gás ou mesmo outros efeitos cosmológicos com a energia escura. Contudo, se f_{NL} de facto variar com a escala, as não-Gaussianidades podem ter uma maior amplitude positiva/negativa à escala dos enxames e portanto os resultados obtidos sugerem que as não-Gaussianidades primodiais devem ser tidas em conta quando dados observacionais de enxames de galáxias são utilizados para determinar parametros cosmológicos ou para inferir o poder de constrangimento de futuras *surveys*.

Por fim, apresentamos o COSMOABC, um código Python, para Computação Bayesiana Aproximada (*Approximate Bayesian Computation* - ABC) baseado no algoritmo *Population Monte Carlo*, que é uma variante do ABC original. O código, que faz uso de um esquema adaptativo de *importance sampling*, é muito flexível, altamente paralelizado e pode ser facilmente interligado com códigos externos, dando ainda possibilidade ao utilizador de definir as suas próprias funções de distância e *priors*. Apresentamos ainda um exemplo de uma aplicação prática, onde acoplamos o COSMOABC com a biblioteca NUMCOSMO e demonstramos com é possível estimar a distribuição posterior de probabilidade de parâmetros cosmológicos, com base em dados observacionais simulados da abundância de enxames de galáxias, sem explicitamente calcular a *likelihood*.

Acknowledgements

First, I wish to thank my supervisors Professor Pedro Viana and Professor Pedro Avelino for their support and extremely helpful comments during the development of my research work, that constituted this thesis. I am and will be forever in debt to Dr. António da Silva for his friendship, endless conversations (sometimes not about cosmology), for educating me on N-body simulations, group-finding algorithms, galaxy clusters scaling relations and for being always available to help me with whatever problem I came across during my PhD. I'm deeply grateful to Ana and Florinda, for the all the moments of love, joy, fun, craziness and for being my safe haven in those moments of darkness and madness of the PhD. A very special thanks to my closest and dear friends Claudo, Odete, Fabio for their endless friendship and motivation. To all the members of the Cosmostatistics Iniciative (COIN). I want to thank all the people of IA at Porto and Lisbon, especially to Lara, Isa, Sérgio and to the IA-Porto football team. I dedicate this thesis to my parents, my brother, to Ana and Florinda, without whom this thesis would not be possible. I acknowledge the financial support of FCT/IDPASC through the grant SFRH/BD/5167/2011.

Preface

Part of the research work presented in this thesis are based on the following publications:

1. **"Is $w \neq -1$ evidence for a dynamical dark energy equation of state?"**
   P. P. Avelino, A. M. M. Trindade, P. T. P. Viana, Physical Review D, 80, 067302
   arXiv:0906.5366
   In this paper, my contribution was the derivation of the algebraic solutions for the evolution of the quintessence scalar field responsible for dark energy. These results are presented in section 1.3 of the thesis.

2. **"A new signature of primordial non-Gaussianities from the abundance of galaxy clusters"**
   A. M. M. Trindade, P. P. Avelino, P. T. P. Viana, Monthly Notices of the Royal Astronomical Society, 424, 1442
   arXiv:1109.6778
   All the numerical work, statistical analysis of the simulations output and the writing of most of the paper was done by me. The results of this paper appear in section 2.3 of chapter 2 of the thesis.

3. **"Biased cosmological parameter estimation with galaxy cluster counts in the presence of primordial non-Gaussianities"**
   A. M. M. Trindade, P. P. Avelino, P. T. P. Viana, Monthly Notices of the Royal Astronomical Society, 435, 782
   arXiv:1304.4275
   The numerical computations, parameter estimations using MCMC, analysis of the simulations output presented in this paper were all done by myself, as well as most of the written text of the paper. The results obtained are shown in section 2.4 of chapter 2.

4. **"cosmoabc: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation"**
   E. E. O. Ishida, S. D. P. Vitenti, M. Penna-Lima, J. Cisewski, R. S. de Souza, A. M. M. Trindade, E. Cameron, V. C. Busti, for the COIN collaboration; Astronomy and Computing, 13, 1
   arXiv:1504.06129
   My contribution to this paper was the implementation of the package in Python and its parallelization. I have also participated discussion and writing of sections in the paper. This work

constitutes chapter 3 of the thesis.

5. **"Effect of Primordial non-Gaussianities on Galaxy Clusters Scaling Relations"**
   A. M. M. Trindade, António da Silva, submitted to Monthly Notices of the Royal Astronomical Society
   arXiv:1603.09270
   I generated all the initial conditions boxes and ran all the the N-body hydrodynamical simulation used in this research paper. I was also responsible to perform all the post-simulation statistical analysis, related plots and wrote some of sections of the paper. The results of this work constitutes section 2.5 of chapter 2 of the thesis.

In the final stages of the PhD also participated in the following research work:

"Exploring the spectroscopic diversity of type Ia supernovae with DRACULA: a machine learning approach "
Michele Sasdelli, E. E. O. Ishida, R. Vilalta, M. Aguena, V. C. Busti, H. Camacho, A. M. M. Trindade, F. Gieseke, R. S. de Souza, Y. T. Fantaye, P. A. Mazzali, for the COIN collaboration; submitted to Monthly Notices of the Royal Astronomical Society
arXiv:1512.06810
My tasks in this publication was the implementation of the self-organizing maps (SOM) method, its integration on the overall DRACULA package and the writing of the respective sections of the paper.

Contents

List of Tables

List of Figures

Introduction

Long before it can be remembered, mankind has always had a particular fascination about the heavens. We have gone from simply wondering about the tiny bright spots in the night sky to building mathematical (complex) models to describe the cosmos. The birth of modern cosmology occurred not long ago and its key building blocks were established throughout the 20^{th} century, benefiting from important theoretical and observational breakthroughs.

In 1915, a revolutionary new theory of gravity, called *General Relativity* (hereafter GR), was proposed by Albert Einstein [1]. In this theory of gravity, space and time are regarded as a single entity and modelled as a four dimensional continuum (known as *spacetime*). The local geometric properties of spacetime are encoded in the so-called *metric*, whose behaviour is affected by all forms of energy. In other words, the metric is determined by the energy-momentum tensor, $T_{\mu\nu}$, and the dynamics of free particles is given by free-fall geodesics. The dynamics of the spacetime is then tracked by the *Einstein field equations*.

It was by the hand of Einstein himself, in 1917, that the first attempt to describe the Universe at large scales, resorting to GR, was made [2]. Assuming that the Universe is homogeneous and isotropic, Einstein proposed a static, time-independent solution of the GR field equations. However, a static solution was not allowed by the original GR equations. To circumvent this problem, Einstein introduced an additional term in his original equations, consisting of a (positive) Cosmological constant (Λ) acting as a repulsive force to counteract the gravitational attraction of ordinary matter. Several years later, the model that Einstein had suggested, was shown to be incorrect, since observations were suggesting that the Universe was in fact expanding. Remarkably, in the same year that Einstein proposed his static model, de Sitter, assuming a vacuum dominated universe (i.e., an universe without matter), had shown that the inclusion of a positive cosmological constant term in the Einstein equations would give rise to a non-static solution. Such solution describes a universe in accelerated (exponential) expansion [3]. Through the years of 1922 to 1924 and later in 1927, Friedmann [4, 5], Lemaître [6], Robertson [7, 8] and Walker [9], found a new homogeneous and isotropic solution of the Einstein's field equations. This solution, known as *Friedmann-Lemaître-Robertson-Walker* (hereafter FLRW) metric, plays a fundamental role in today's mathematical description of the Universe.

These mathematical models were the first attempts to provide a meaningful and consistent description of the Universe, at a time when observations at large cosmological scales were almost non-existent. In fact, it was not until 1929 that a major observational achievement, known today as the *Hubble Law*, was established [10]. By observing distant faint galaxies, Edwin Hubble showed

that they were receding from Earth, with velocities proportional to their distances, thus indicating that the Universe was indeed expanding and invalidating the static model proposed by Einstein.

The relentless search for a more complete mathematical description of the Universe led to the formulation in the 1960s of the Standard Cosmological Model (hereafter SCM), which allows us to quantitatively understand the Universe, from a fraction of a second old ($t \approx 10^{-2} - 10^{2}$ sec) from the big-bang, up to the present time ($t \approx 14$ Gyrs). The SCM "paints" a picture of the Universe as starting from a primordial hot and dense initial state at some point in the far past and being expanding ever since.

A remarkable prediction of the SCM is the existence of a residual black body radiation resulting from an epoch when matter and radiation were in thermodynamic equilibrium. As pointed out by Gamow and his collaborators [11–13], such radiation should still permeate the Universe in the form of a thermal microwave background with a predicted temperature around 5 Kelvin[1] [14]. However, such background radiation, now known as *Cosmic Microwave Background Radiation* (hereafter CMB), remained elusive from observational evidence for many years. In 1965, a very important discovery was made by accident by Penzias and Wilson [15], at the time two employees of the Bell Telephone Laboratories, New Jersey. They picked up a mysterious microwave static signal on their horn antenna, which seemed to be equally distributed across the sky and that had a temperature comparable to the one predicted by Gamow. Soon after, Dicke and collaborators [16] interpreted this "lucky" accident as the CMB, the relic radiation from the primordial Universe.

The high level of homogeneity and isotropy of the CMB at large scales is probably the most striking observational evidence in favour of the Cosmological Principle. However, at smaller scales, the Universe is populated with galaxy clusters, superclusters, voids, etc, thus not being perfectly homogeneous and isotropic. In fact, little time had passed since the discovery of the CMB radiation, when it was conjectured that the initial density perturbations that seeded the structures we observe today, would also have left their "fingerprints" in the background radiation in the form of primary temperature anisotropies. Those would manifest themselves in the CMB temperature power spectrum as a slight differences in temperature in different directions of observation. Due to the small amplitude of these anisotropies, its detection would be extremely difficult (see [17, 18]) and it would constitute another important test of the SCM. The first detection of these anisotropies was done with the *Differential Microwave Radiometers* (DMR) instrument on board of the *Cosmic Background Explorer* (COBE) satellite, by Smoot and collaborators in 1992 [19]. Since then, more sophisticated and precise instruments on board of NASA's *Wilkinson Microwave Anisotropy Probe* (WMAP) in 2001 and ESA's Planck satellite in 2009 have enabled us to measure and characterize primordial anisotropies with incredible accuracy. Figure 1 shows the CMB temperature anisotropy maps obtained by COBE, WMAP and PLANCK satellites. Bluer spots/regions correspond lower temperatures, while redder spots/regions have higher temperatures.

Since the last decade, our ability to study of the Universe with high accuracy has suffered tremendous improvements, benefiting not only from the huge quantity of high quality observational data available, but also from the very fast growth in computational power and cutting-edge statistical analysis techniques. In fact, just a century ago, images taken with photographic plates in

[1] Current observations of the Planck satellite set this temperature around 2.73 Kelvin.

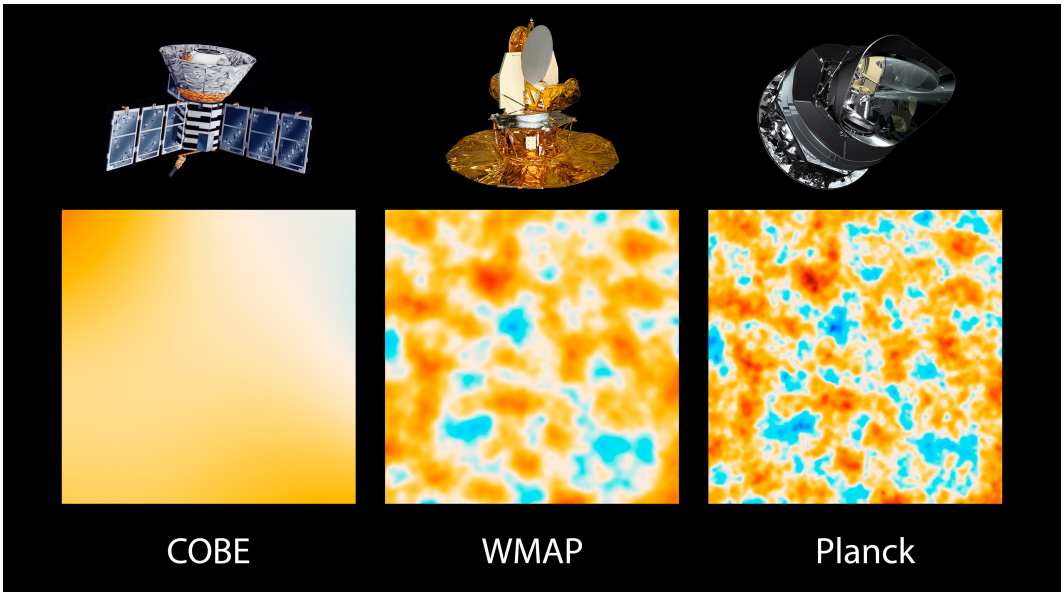
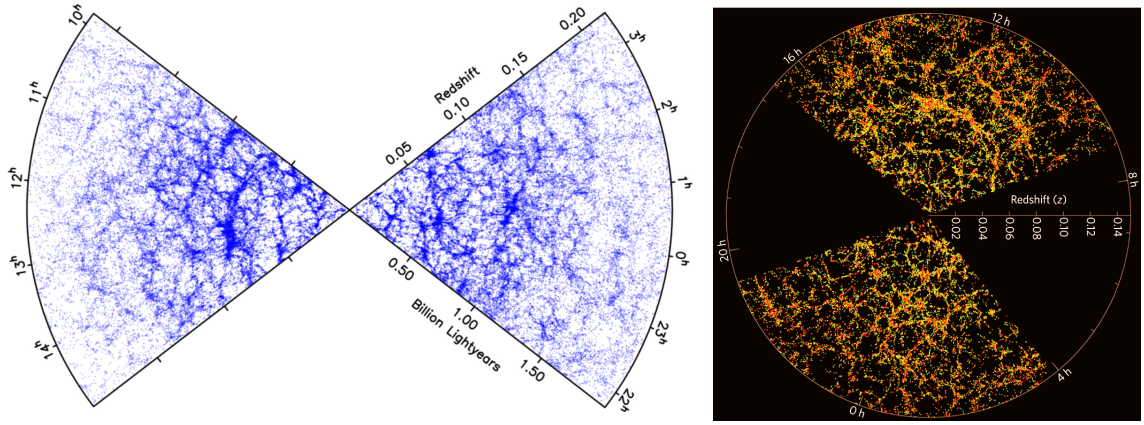


Figure 1: The comparison of the CMB temperature measurements made by COBE (left), WMAP (centre) and Planck (right). The image shows tiny small variations (anisotropies) in the temperature readings (more evident for WMAP and Planck). Hotter (colder) regions of the sky are represented by red (blue). Image credit: NASA/JPL-Caltech/ESA.

small aperture telescopes were the only open window to the Cosmos. Today, rapid advances in telescope and detector technology, both in space and on the ground, have enabled the observation of the Cosmos spanning from the radio waves to the gamma ray band of the electromagnetic spectrum. These observations have provided us a rough picture of the processes that have taken place throughout the evolution of the Cosmos, from its simpler hot beginning to the more complex present. Although the SCM gives an accurate description of a wide range of observations with a minimum number of free parameters, it does not provide answers to some puzzling questions about the origin and evolution of the Universe. For instance, observations indicate that only a small percentage of the energy content of Universe is composed of ordinary matter. Roughly a quarter is dark matter, a component that manifests itself essentially through the gravitational interaction. However, maybe the most striking observational result is that the predominant component of the Cosmos is Dark Energy (hereafter DE), a mysterious entity with gravitationally repulsive properties and whose existence was postulated in order to explain the current accelerated expansion of the Universe, as suggested by Type Ia supernova data (see [20, 21]). Nowadays, understanding DE, what is its underlying fundamental nature, how it affected, affects and will affect the dynamics of the Universe, is one of the most challenging and highly active areas of research within Modern Cosmology.

Another puzzling issue in Modern Cosmology is the formation of the Large Scale Structure (hereafter LSS) of the Universe. Large galaxy surveys such as the Two-degree-Field Galaxy Redshift Survey [22] (2dFGRS) and the Sloan Digital Sky Survey [23] (SDSS), have shown that galaxies and mass are not uniformly distributed throughout the Cosmos. Instead, they exhibit a web-like structure - known as “Cosmic Web”, consisting of dense compact clusters, walls, elongated filaments and large voids (see figure 2).



(a) (b)

Figure 2: The spatial distribution of galaxies obtained by the 2dfFGRS [24] (a) and SDSS [25](b) surveys.

The origin of the primordial density inhomogeneities that gave rise to the LSS in the Universe is also source of great debate within cosmology. These structures are thought to originate from the late-time process of gravitational collapse of tiny perturbations on the primordial matter density field. The SCM postulates the existence of such primordial fluctuations rather than providing a mechanism from where these fluctuations naturally occur. As it will be discussed later on in this thesis, the inflationary paradigm, originally drafted by Alan Guth in 1981 [26], provides a simple, yet powerful, mechanism to generate the seeds of LSS. Inflation refers to a period of accelerated expansion at early stages of the Universe and within this paradigm primordial density fluctuations may arise from quantum fluctuations of the inflaton field responsible for driving inflation.

One of the most fundamental predictions of the simplest standard, single field, slow-roll inflationary model, is that the initial density perturbations, that seeded the formation of the LSS we see today, were generated randomly, with a Gaussian probability distribution associated with their amplitudes (see e.g. [27–31]). Although such prediction seems to be in good agreement with current observations of the cosmic microwave background anisotropies [32–35] and LSS data (e.g. [36]), a significant and potentially observable level of non-Gaussianity may be produced in inflationary models in which any of the conditions associated to the standard single-field, slow-roll inflation fail. Detecting and constraining primordial non-Gaussianities has become a crucial task in current cosmological studies, since a positive detection of non-Gaussianity would rule out a considerable number of inflationary models, opening an entirely new window into the very early Universe.

A wide range of observables has been used to probe primordial non-Gaussianities. Traditionally, the CMB temperature anisotropies three-point statistics has been the tool of choice to try to detect and constrain deviations from Gaussianity in the primordial matter density field. However, this task could also be achieved through the statistical characterization of the properties of the large-scale structure. Namely the bispectrum and/or trispectrum of the galaxy distribution (e.g. [31, 37, 38]) and weak-lensing observations (e.g. [39, 40]), as well as CMB-LSS [41] and CMB-21cm [42] cross-correlations have been used to the same effect. The evolution with time of the abundance of massive collapsed objects such as galaxy clusters (e.g. [43–49]) also holds key

information that could be used to probe primordial non-Gaussianities. These objects are the most massive gravitationally bounded structures in the Universe and they form at the high peaks of the matter density field. Their number density as a function of redshift depends on the growth of structure, thus being sensitive to the dynamics and energy content of the Universe and to the statistical properties of the primordial matter density fluctuations. Large voids has also been presented in the literature as another method to constrain cosmology and in particular non-Gaussian models (see e.g [50–53]).

If indeed, theoretical models that allow deviations from Gaussianity in the initial density field are correct, then assumptions of Gaussianity will inevitable lead to biases in the estimation of cosmological parameters from observations. The central scientific objective of this thesis is the characterization of the biases that may be introduced in the estimation of cosmological parameters, as a result of assuming that the probability distribution associated with the amplitude of density perturbations is Gaussian. Particular attention is given to biases that may arise in the estimation of the amplitude of the matter density fluctuations, the density of matter and the density and equation of state of the dark energy. The methods that will be considered will be those associated with the growth of large-scale structure in the Universe, like the evolution with time of the number density of collapsed objects, such as galaxy clusters. Moreover, we also investigate the effects of primordial non-Gaussianities on galaxy clusters scaling relations. This is done resorting to a set of N-body hydrodynamical numerical simulations of large scale structure with gas physics, allowing for different levels of non-Gaussian initial conditions.

A second objective of the thesis is the development of algorithms and techniques for the extraction of key information from observational data, eyeing future observing facilities like the already approved Large Synoptic Survey Telescope (LSST), or the proposed EUCLID mission (ESA Cosmic Vision 2015-2025) and Square Kilometre Array (SKA). On this front, we introduce and motivate a novel package called COSMOABC, which was developed within the CosmoStatistics Iniciative (COIN) Collaboration. This package is a Python Approximate Bayesian Computation (herafter ABC) sampler featuring a *Population Monte Carlo* variation of the original ABC algorithm, which uses an adaptive importance sampling scheme. We also show an example of practical application of this package of how it can be used to estimate the posterior probability distribution of cosmological parameters, based on measurements of galaxy clusters number counts without computing the likelihood function.

This thesis is organized as follows. In chapter 1 we describe a few basic but important concepts such as the standard cosmological model, dark energy, the inflationary paradigm and structure formation. In chapter 2, we briefly study primordial non-Gaussianities, how they arise in many inflationary models. Furthermore, we investigate their impact both on the determination of cosmological parameters and on the characterization of galaxy clusters scaling laws. In chapter 3 we introduce and motivate a novel python package, the COSMOABC, to perform Bayesian parameter estimation in cosmology. A practical cosmological application of this package is also done.

Chapter 1

The Standard Cosmological Model - Review

In this chapter we briefly review some key ingredients of modern cosmology, such as the Standard Cosmological Model, the Inflationary paradigm and Structure Formation theory, which are the basic foundations of this thesis. The great majority of the results shown here can be easily found in many of the usual cosmology textbooks [54–62] and we refer to them for more detailed derivations of the results that will be discussed in the following sections.

1.1 The Homogeneous and isotropic Universe

The Standard Cosmological Model is the widely accepted physical description of the Universe and its evolution. It relies on two basic pillars. The first is that the dynamical properties of the Universe can be described resorting to Einstein's theory of General Relativity. The second is that, on large scales, the Universe is homogeneous and isotropic. The second of these pillars is a direct consequence of the Cosmological Principle, which states that the properties of the Universe are the same everywhere in space and in every direction. Although this is true on very large scales (scales of the order of hundreds of Mpc), on smaller scales stars, galaxies, galaxy clusters, voids and other structures populate the Universe, making it highly inhomogeneous and anisotropic.

Assuming that the Cosmological Principle holds, a solution of the Einstein equations of GR is described by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, whose line element can be written as,

$$ds^2 = -dt^2 + a^2(t)\left[d\chi^2 + F_K^2(\chi)\left(d\theta^2 + \sin^2(\theta)\, d\phi^2\right)\right], \tag{1.1}$$

where t is the cosmic time measured by a comoving observer, χ is the comoving radial distance, θ and ϕ are angular coordinates, and $a(t)$ is the so-called scale factor which encodes the dynamics

of the Universe. The function $F_K(\chi)$ is given by,

$$F_K(\chi) = \begin{cases} \sinh\chi, & K=-1 \\ \chi, & K=0 \\ \sin\chi, & K=+1 \end{cases} \tag{1.2}$$

where the spatial curvature parameter $K = -1, 0, 1$ defines a open, flat and closed universe, respectively.

The scale factor enables us to track and study the dynamics of the expanding Universe. It describes the evolution of the proper distance between two objects or observers in space moving with the Hubble flow at a given time t. It can also be expressed in terms of the redshift z. To see this, let us consider two observers A and B at some comoving distance from each other in an expanding universe. At some time t_1, the observer A emits a photon, with wavelength λ_1, that is detected by the observer B at some time t_2, with wavelength λ_2. Due to the expansion, the spectrum of the emitted light will acquire a redshift, which is defined as the change in the energy of the photon between the emission time t_1 and the detection time t_2. Mathematically, the redshift is written as,

$$z = \frac{E_1 - E_2}{E_2} = \frac{\lambda_2}{\lambda_1} - 1 = \frac{a(t_2)}{a(t_1)} - 1. \tag{1.3}$$

In eq. (1.3), E and λ are respectively, the photon's energy and wavelength. If the observer B is detecting the photon at the present time, t_0, then we obtain,

$$z = a^{-1} - 1, \tag{1.4}$$

where we have set $a(t_0) = a_0 = 1$.

It is also possible to define the proper (physical distance) between comoving observers A and B as

$$d_p(t) = a(t)\,\chi. \tag{1.5}$$

From the above equation we define the relative speed between both observers by performing a derivative with respect to cosmic time,

$$v(t) = \frac{d}{dt} d_p(t) = \frac{da(t)}{dt}\chi = \frac{\dot{a}(t)}{a} d_p \equiv H(t)\, d_p(t). \tag{1.6}$$

If the Universe is expanding, i.e. $\dot{a} > 0$, the previous expression tells us that the larger the distance between two observers, the more rapid they recede from each other. Eq. (1.6) is the well known Hubble law and it was inferred from observations for the first time by Edwin Hubble in 1929 [10]. The expansion rate of the Universe, at any given time, is hidden in the proportionally factor of eq. (1.6) and its current value is usually written as $H_0 = 100h$ km/s/Mpc. H_0 has units of inverse time and it is known as the Hubble parameter. Current observational constraints on the value of this parameter are $H_0 = (67.4 \pm 1.4)$ km/s/Mpc [14].

1.2 Friedmann Equations

The Theory of General Relativity, developed by Albert Einstein (1905 to 1915) sets the underlying building block of the Standard Model of Cosmology. Eintein's field equations describe how a distribution of matter, described by the energy-momentum tensor $T_{\mu\nu}$, influences the four-dimensional spacetime geometry and vice-versa. These equations are given by,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}, \tag{1.7}$$

where G is the gravitational constant, g is the determinant of the metric $g_{\mu\nu}$ whose signature we assume here to be $(-,+,+,+)$, R and $R_{\mu\nu}$ are the Ricci scalar and the Ricci tensor, respectively.

To solve the Einstein field equations (eq. (1.7)) based on the assumption that the Universe is homogeneous and isotropic (again, assuming that the Cosmological Principle holds), the energy-momentum tensor has also to be consistent with these assumptions. Here we assume $T_{\mu\nu}$ is that of a perfect fluid, given by

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p g_{\mu\nu}, \tag{1.8}$$

In eq. (1.8), u_μ is the four-velocity of the fluid, and ρ and p are the density and pressure of the fluid, respectively. The condition of homogeneity implies that both ρ and p are only time dependent quantities. On the other hand, isotropy implies that the four-velocity has only a time component, i.e. $u_\mu = (1,0,0,0)$. Therefore, in comoving coordinates the energy-momentum tensor is diagonal and given by,

$$T_{\mu\nu} = diag\,(\rho, p, p, p)\,. \tag{1.9}$$

Inserting the FLRW metric (eq. (1.1)) and the previous energy-momentum tensor (eq. (1.8)), in the Einstein field equations (eq. (1.7)) we obtain, after some algebraic manipulation, the Friedmann and Raychaudhuri equations, given respectively by

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}, \tag{1.10}$$

$$\left(\frac{\ddot{a}}{a}\right) = -\frac{4\pi G}{3}\left(\rho + 3p\right). \tag{1.11}$$

The conservation law of the energy-momentum tensor is expressed by the condition

$$T^{\mu\nu}{}_{;\nu} = 0, \tag{1.12}$$

where ";" denotes the covariant derivative. The temporal component of eq. (1.12) corresponds to the continuity equation for the energy density,

$$\frac{d\rho}{dt} + 3H\left(\rho + p\right) = 0, \tag{1.13}$$

which can also be written in the form

$$\frac{d}{dt}\left(\rho a^3\right) + p\frac{d}{dt}a^3 = 0. \tag{1.14}$$

Having in mind the first law of thermodynamics of a fluid in equilibrium is $TdS = dU + pdV$, where T and S are respectively the temperature and entropy; then eq. (1.14) describes an adiabatic expansion of an homogeneous and isotropic perfect fluid element with comoving volume $V \propto a^3$ and energy $U \propto \rho a^3$.

Before solving the Friedmann equations to determine the scale factor, one more ingredient is needed: we must specify the equation of state that relates the pressure and the energy density of the perfect fluid. In many cases of interest in cosmology, pressure and density are related via the expression,

$$p = w\rho. \tag{1.15}$$

In eq. (1.15), w is the fluid's equation of state parameter and its value depends on the type and specific properties of the fluid/fluids filling the universe. Further assuming that w is constant over time, a generic solution for the continuity equation is given as,

$$\rho a^{3(1+w)} = \rho_0, \tag{1.16}$$

where ρ_0 is the present value of the density and we have set $a_0 = 1$ as before. The cases $w = -1$, $w = 0$, $w = 1/3$ are classical examples which adequately describe the important epochs of the Universe dominated respectively by a fluid of constant vacuum energy, non relativistic matter and radiation. However, if the equation of state varies with time, i.e. $w = w\left(a\left(t\right)\right)$, the solution for the continuity equation (eq. (1.13)) acquires the general form,

$$\rho = \rho_0 \exp\left(-3\int_a^1 d\ln a\left[1 + w\left(a\right)\right]\right). \tag{1.17}$$

From the Friedmann equations it is also possible to define, at any given time, the critical density of the Universe, $\rho_c\left(t\right) = 3H^2/8\pi G$, which is the density for which the Universe is perfectly flat ($k = 0$) and so eq. (1.10) may be re-written in the form,

$$\Omega\left(t\right) \equiv \frac{\rho}{\rho_c} = -1 + \frac{K}{\left(aH\right)^2}. \tag{1.18}$$

The above equation implies that the energy content of the universe determines its spatial geometry, in other words we have,

$$\Omega < 1 \text{ or } \rho < \rho_c \quad \Rightarrow K = -1 \tag{1.19}$$

$$\Omega = 1 \text{ or } \rho = \rho_c \quad \Rightarrow K = 0 \tag{1.20}$$

$$\Omega > 1 \text{ or } \rho > \rho_c \quad \Rightarrow K = 1 \tag{1.21}$$

It is commonly assumed that the Universe is made of a mixture of three perfect fluids, each one contributing to the Universe's energy density: pressureless matter ρ_m, radiation ρ_r and a cosmological constant ρ_Λ. Using eqs. (1.17) and the definition of critical density, we may write eq. (1.10) as,

$$H^2 = H_0^2 \left(\Omega_{m0} a^{-3} + \Omega_{r0} a^{-4} + \Omega_{\Lambda 0} + \Omega_{K0} a^{-2}\right), \tag{1.22}$$

where we have set $w = 0$, $w = 1/3$, $w = -1$, for matter, radiation and cosmological constant, respectively. Furthermore, the last term on the right side of eq. (1.18) refers to the curvature parameter, defined as,

$$\Omega_K \equiv -\frac{K}{a^2 H^2}. \tag{1.23}$$

The quantities H_0, Ω_{m0}, Ω_{r0} and $\Omega_{\Lambda 0}$ are respectively the Hubble parameter (defined previously), and the fractional energy density of matter, radiation and cosmological constant at present time. A measure of the rate of acceleration of Universe is often done by resorting to the *deceleration parameter*, which is define as,

$$q \equiv -\frac{\ddot{a} a}{\dot{a}^2}. \tag{1.24}$$

The expansion rate is initially dominated by the radiation fluid term, Ω_{r0}. As the expansion continues, the contribution from the other terms start to become more and more important. Eventually, the Universe suffers a transition from a radiation dominated epoch to a matter dominated one and, later on, the dynamics becomes dominated by the cosmological constant term. From the first two terms of eq. (1.22) it is possible to compute the redshift z_{eq} for which the Universe changes from radiation to matter dominance,

$$1 + z_{eq} = \frac{\Omega_{m0}}{\Omega_{r0}}. \tag{1.25}$$

Similarly, the epoch at which both matter and cosmological constant contribute equally to the expansion rate is given by,

$$1 + z_\Lambda = \left(\frac{\Omega_{\Lambda 0}}{\Omega_{m0}}\right)^{1/3}. \tag{1.26}$$

From [14] the constraints on the previous quantities are $z_{eq} = 3386 \pm 69$, $\Omega_{m0} = 0.314 \pm 0.020$, $\Omega_{\Lambda 0} = 0.686 \pm 0.020$. Introducing the current observational values of Ω_{m0} and $\Omega_{\Lambda 0}$ in eq. (1.26), it is possible to see that the transition from matter to cosmological constant dominance in the Universe occurs for a redshift around $z_\Lambda \sim 0.3$.

1.3 Dark Energy

Since the discovery made by Hubble in 1929 that the Universe was expanding, cosmologists have tried to measure the slowing down of the cosmic expansion due to gravity, as it was expected for a universe dominated by baryonic matter and radiation. However, in 1998 the results obtained by two

independent teams studying distant Type Ia supernovae [20, 21] surprised most of the cosmology community. Their results showed that, contrary to what was expected, the Universe's expansion rate was accelerating. However, the precise origin of this cosmic acceleration remains a deep mystery. One possibility is that General Relativity breaks down at cosmological scales and thus a more complete theory of gravity is required (see [63] and references therein). However, assuming that GR holds, from eq. (1.11), an accelerated expansion is only possible if $\rho + 3p < 0$. If the Universe is constituted only by ordinary matter and radiation, this condition is not fulfilled, since gravity would slow down the expansion. Hence, in order to reconcile theory and observations, a new contribution to the energy-momentum tensor had to be proposed. Dark Energy, an exotic fluid with negative pressure and contributing nearly three-quarters of the energy content of the Universe, is the most widely accepted hypothesis to explain the cosmic accelerated expansion. The precise nature of Dark Energy has not yet been uncovered and constitutes one of the most important open problems in Modern Cosmology.

The Cosmological Constant Analysing the Friedmann equation more in depth we realize that if $p > -\rho/3$ then $\ddot{a}(t)$ is always negative ($\ddot{a}(t) < 0$), which corresponds to a decelerated expansion of the universe. However, when Einstein formulated his theory of gravity it was thought that the Universe was static and immutable. Thus, in order to obtain a static cosmological model, Einstein introduced a new term proportional to the metric in the left-hand side of eq. (1.7), the cosmological constant Λ,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{1.27}$$

As a consequence, the Friedmann equations must also be modified. This can be achieved by replacing ρ and p by

$$\rho \to \rho + \frac{\Lambda}{8\pi G}, \tag{1.28}$$

$$p \to p - \frac{\Lambda}{8\pi G}. \tag{1.29}$$

It is important to note that Einstein interpreted the need for a cosmological constant as a modification of the equations on the geometry side. Nowadays, it is thought to be an intrinsic, fundamental energy of the quantum vacuum [64] contributing to the energy-momentum tensor and accounting roughly for more than two-thirds of the total energy density of the Universe.

The Cosmological Constant (hereafter Λ) is the most simple Dark Energy candidate and can be described as a spatially homogeneous fluid whose equation of state parameter is constant over time and given by $w_\Lambda = \frac{P_\Lambda}{\rho_\Lambda} = -1$ (where $P_\Lambda$and ρ_Λ are the pressure and energy density respectively). It is thought to be associated with the energy density of the vacuum, having an energy-momentum tensor of the form $T_{\mu\nu} = -\rho_{vac} g_{\mu\nu}$. Quantum mechanics predicts a non-zero ground-state energy for the vacuum, which is due to spontaneous creation and annihilation of particles and anti-particles around the "zero point" energy state. The value of the energy density

of the vacuum has been computed theoretically using quantum field theory and it was estimated to be $\rho_{\Lambda 0} \sim 10^{92}$ erg/cm^3 [65]. Normalizing the previous value by the critical density, then one gets $\Omega_{\Lambda 0} \sim 10^{120}$, which strongly deviates from the present value of $\Omega_\Lambda = 0.686 \pm 0.020$ estimated from observational data [14].

Dynamical Dark Energy Numerous other candidates have been proposed in the literature. Much discussed have been scenarios where the dark energy sector is modelled by a minimally coupled homogeneous scalar field, named quintessence, which slowly rolls down a runaway potential.The action for a generic scalar field can be written as

$$S = \int d^4x\sqrt{-g}\,\mathcal{L}\,(\phi, X)\,, \tag{1.30}$$

where $\mathcal{L}$ is Lagrangian density , $X = -\frac{1}{2}\phi_{,\mu}\phi^{,\mu}$ and a comma is used to represent a partial derivative. The energy-momentum tensor of the scalar field may be written in a perfect fluid form

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu}\,, \tag{1.31}$$

by means of the following identifications

$$u_\mu = \frac{\phi_{,\mu}}{\sqrt{2X}}\,, \quad \rho = 2X\mathcal{L}_{,X} - \mathcal{L}\,, \quad p = \mathcal{L}(X,\phi)\,. \tag{1.32}$$

In eq. (1.31), $u_\mu = (1,0,0,0)$ is the normalized 4-velocity field describing the motion of the fluid (for timelike $\phi_{,\mu}$), while ρ and p are its proper energy density and pressure, respectively. The dark energy equation of state parameter w is

$$w \equiv \frac{p}{\rho} = \frac{\mathcal{L}}{2X\mathcal{L}_{,X} - \mathcal{L}}, \tag{1.33}$$

and the sound speed squared is given by

$$c_s^2 \equiv \frac{p_{,X}}{\rho_{,X}} = \frac{\mathcal{L}_{,X}}{\mathcal{L}_{,X} + 2X\mathcal{L}_{,XX}}\,, \tag{1.34}$$

as long as $\mathcal{L}_{,X} \neq 0$.

For the particular case of quintessence dark energy models, described by a homogeneous real scalar field, the Lagrangian density is given by,

$$\mathcal{L} = X - V(\phi)\,. \tag{1.35}$$

Generically, the equation describing the dynamics of a scalar field may be obtained by varying the action with respect to ϕ

$$\frac{1}{\sqrt{-g}}\left(\sqrt{-g}\mathcal{L}_{,X}\phi^{,\mu}\right)_{,\mu} = \mathcal{L}_{,\phi}\,. \tag{1.36}$$

Assuming a flat Friedmann-Robertson-Walker metric eq.(1.1), the scalar field equation of motion

is approximately given by

$$\frac{\partial}{\partial t}\left(\mathcal{L}_{,X}\frac{\partial\phi}{\partial t}\right)+3H\mathcal{L}_{,X}\frac{\partial\phi}{\partial t}=\mathcal{L}_{,\phi}, \tag{1.37}$$

which, by introduction of the proposed Lagrangian, reduces to

$$\ddot{\phi}+3H\dot{\phi}+V_{,\phi}=0\,, \tag{1.38}$$

where a dot represents a derivative with respect to the physical time, t. The dark energy equation of state parameter w is given by

$$w\equiv\frac{p}{\rho}=\frac{\dot{\phi}^2/2-V(\phi)}{\dot{\phi}^2/2+V(\phi)}\,, \tag{1.39}$$

and the sound speed squared is $c_s^2=1$. The fact that the sound speed is equal to the speed of light prevents the generation of large spatial fluctuations in the dark energy density. From eq. (1.39), it is possible to see that if the scalar field evolves slowly, i.e., if $\dot{\phi}^2/2\ll V$, then $w\approx-1$ and the field behaves like a cosmological constant, with $\rho_\phi\approx V(\phi)$.

The functional form and time evolution of the equation of state parameter will depend on the shape of the dark energy potential $V(\phi)$. Due to the large variety of models available in the literature, the task of confronting all of them directly with observations is not practical. A way to overcome this, is to define a parametrization for the equation of state parameter. The most obvious one other than $w=-1$ is to consider that w is time independent, i.e. $w(t)=w=$ constant. The best observational constrains on a constant w are given by [14], which have set its value using a combination of cosmological probes:

$$w=-1.13^{+0.24}_{-0.25}\quad(95\%;\,\text{Planck}+\text{WP}+\text{BAO})\,, \tag{1.40}$$

$$w=-1.09\pm0.17\quad(95\%;\,\text{Planck}+\text{WP}+\text{Union2.1})\,, \tag{1.41}$$

$$w=-1.13^{+0.13}_{-0.14}\quad(95\%;\,\text{Planck}+\text{WP}+\text{SNLS})\,, \tag{1.42}$$

$$w=-1.24^{+0.18}_{-0.19}\quad(95\%;\,\text{Planck}+\text{WP}+\text{H}_0)\,. \tag{1.43}$$

Note that the last constraint is in tension at more than 2σ with $w=-1$.

However, there is no theoretical and/or observational reason, other than mathematical simplicity, why the equation of state should be constant over time. In fact, it has been shown in [66] that if $|w{+}1|\gtrsim0.01$ around the present time, but the dark energy dynamics is sufficiently slow, it is possible that future constraints will rule out a cosmological constant while being consistent with a time-independent equation of state parameter. Although models with such behaviour can be constructed, they do require significant fine-tuning. To see this, let us assume w is constant with time. Hence, eq. (1.39) implies,

$$\ddot{\phi}=V_{,\phi}\frac{1+w}{1-w}\,, \tag{1.44}$$

and

$$\dot{\phi} = \pm \left(2V \frac{1+w}{1-w} \right)^{1/2} . \tag{1.45}$$

From the above equation, we get two possible solutions. In the following we shall drop the $\pm$ sign and consider solutions with $\dot{\phi} > 0$. Thus, substituting eqs. (1.44) and (1.45) into eq. (1.38) one obtains

$$\frac{(V_{,\phi})^2}{V} = \frac{9}{2}(1-w^2)H^2 . \tag{1.46}$$

For this equation to be verified the potential of ϕ would have to be designed such that $V_{,\phi}^2/V \propto H^2$. This requires very large fine-tuning, in particular during the transition from the matter to the dark energy dominated eras. Thus, multiplying equation (1.46) by $\dot{\phi}^2$ and using equation (1.45) it is simple to show that

$$V = V_0 a^{-3(1+w)} , \tag{1.47}$$

where the subscript '0' refers to the present time (we are taking $a_0 = 1$). However, the evolution of ϕ with the scale factor a is, in general, very different in the matter and dark energy dominated eras. In fact, assuming a flat universe, one has

$$H^2 = H_0^2 \left(\Omega_{m0} a^{-3} + \Omega_{de0} a^{-3(w+1)} \right) , \tag{1.48}$$

so that, using eq. (1.45), one obtains

$$\frac{d\phi}{da} = \sqrt{3\Omega_{de0}(1+w)} \left(\Omega_{m0} a^{3w+2} + \Omega_{de0} a^2 \right)^{-1/2} , \tag{1.49}$$

which has the solution

$$\phi = A + B \ln \left(\frac{a^{3w}}{\left(1 + (\Omega_{m0} a^{3w} / \Omega_{de0} + 1)^{1/2} \right)^2} \right) , \tag{1.50}$$

where $B = \sqrt{3(1+w)}/(3w)$, A is an arbitrary integration constant, $\Omega_{m0} = \rho_{m0}/(3H_0^2)$ and $\Omega_{de0} = \rho_{de0}/(3H_0^2)$ is the present day fractional density of dark energy. At very late times ($a \gg 1$) the dark energy will completely dominate the energy density of the universe and the evolution of ϕ with the scale factor will be given by

$$\phi = C_{de} + \sqrt{3(1+w)} \ln a , \tag{1.51}$$

where C_{de} is an arbitrary constant. Using eq. (1.47) one obtains the following solution

$$V = V_{de} \exp \left(-\sqrt{3(1+w)}(\phi - \phi_e) \right) , \tag{1.52}$$

valid at an arbitrary time t_{de} well into the dark energy dominated era (to which the subscript 'de' refers to).

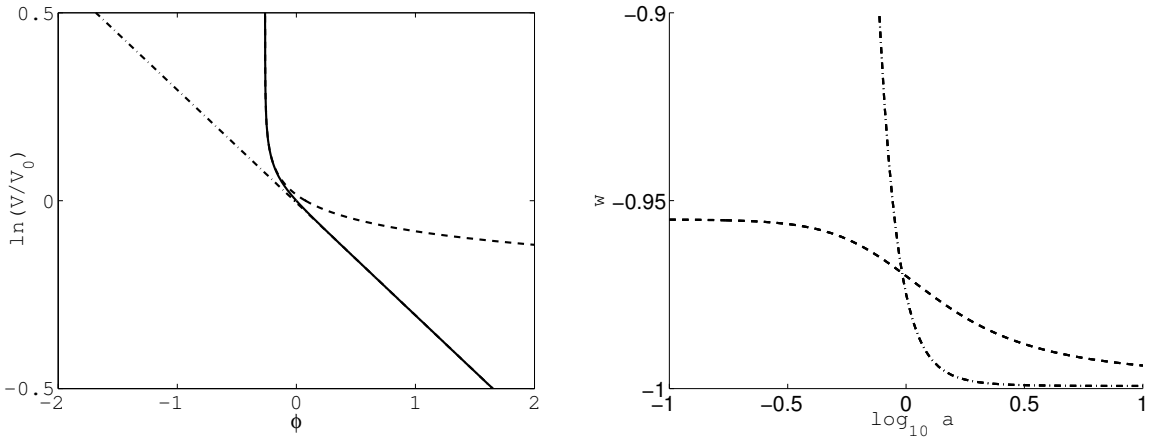


Figure 1.1: a) The solution for $V(\phi)$ assuming that $w = -0.97$ at all times (solid line), as well as the analytical solutions for the scalar field potential, computed using eqs. (1.54) or (1.52), valid deep into the matter era (dashed line) and dark energy era (dot-dashed line) respectively. b) The evolution of the equation of state parameter computed with the potentials given in eqs. (1.54) or (1.52) (dashed and dot-dashed lines, respectively).

(a) (b)

On the other hand, at early times ($a \ll 1$) deep into the matter era one has

$$\phi = C_m - \frac{2}{3w}\sqrt{\frac{3(1+w)\Omega_{de0}}{\Omega_{m0}}}a^{-3w/2}\,, \tag{1.53}$$

where C_m is an arbitrary integration constant. Using eq. (1.47) one obtains the following solution

$$V \propto (\phi - C_m)^{2(w+1)/w}\,, \tag{1.54}$$

valid deep into the matter era.

In figure 1.1 panel (a), we plot the solution for $V(\phi)$ assuming $\Omega_{m0} = 0.27$, $\Omega_{de0} = 0.73$ (taken from the five-year WMAP results [67]) and $w = -0.97$ at all times (solid line), as well as the analytical solutions, computed using eqs. (1.54) or (1.52), valid deep into the matter and dark energy eras (dashed and dot-dashed lines, respectively). The initial conditions for the $w =$ constant solution were chosen so that $\phi_0 = 0$ and the constants C_m and C_{de} were determined by requiring that the analytical solutions computed using eqs. (1.54) or (1.52) fitted the constant w results obtained deep into the matter and dark energy dominated eras, respectively. It is clear from figure 1.1 panel (a) that, in order to obtain $w =$ constant, the shape of the potential must be fine-tuned around $\phi = \phi_0$. Otherwise, the equation of state parameter would change rapidly around the present time. This can be seen in figure 1.1 panel (b), where we plot the evolution of the equation of state parameter with the potentials given by eqs. (1.54) or (1.52) (dashed and dot-dashed lines, respectively). These potentials, designed to produce a constant w deep into the matter and dark energy dominated eras, respectively, give rise to a rapidly changing w in the transition between them, with $|w_0 - w(z=1)|/|w_0 + 1| \gtrsim 1$ (here $z = 1/a - 1$ is the redshift).

The above results show that in order for the dark energy equation of state parameter, w, to be constant in time and close to, albeit different from, -1, a significant amount of fine-tuning would be required in the wide range of models considered. This is essentially the result of the existence of a transition era between matter and dark energy domination in the recent past. Therefore, any future evidence which excludes $w \neq -1$, even if it is consistent with a time-independent value for w, should be interpreted as indicative of a dynamical dark energy equation of state and clearly, in

that situation, a further tightening of the constraints on the time variation of w should be actively sought.

The previous results thus motivates the study of the equation of state in terms of suitable parametrizations which allow for a moderate evolution with the redshift. Many possibilities have been proposed in literature, however the most studied is the parametrization give by [68, 69],

$$w\,(z) = w_0 + w_a(1-a), \tag{1.55}$$

which have been shown to be a robust way to describe a wide diversity of dark energy models. The parameter w_0 is the present value of the equation of state and w_a parametrizes the variation of the equation of state with time.

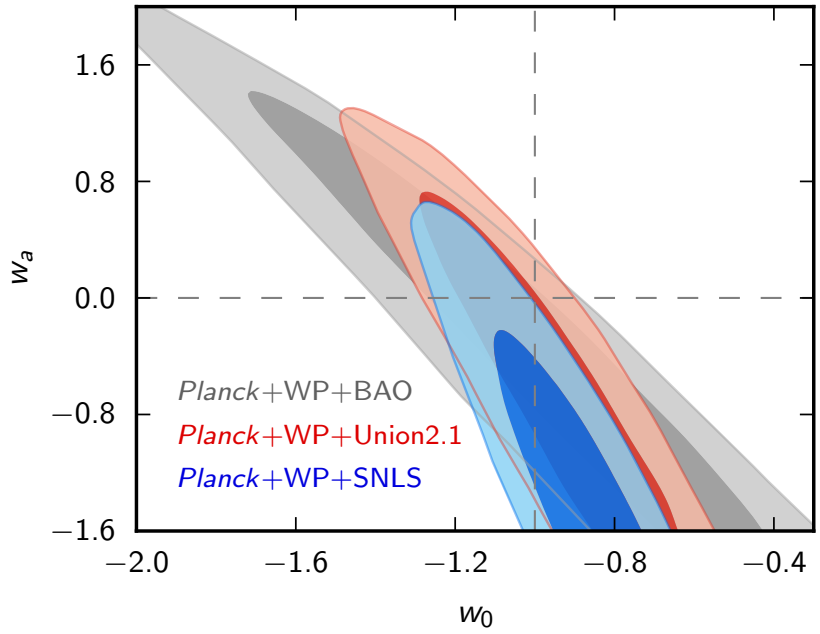


Figure 1.2: The plot shows the marginalized posterior distributions for w_0 and w_a, for the data combinations Planck+WP+BAO (grey), Planck+WP+Union2.1 (red) and Planck+WP+SNLS (blue); as well as the 68% and 95% contours [14]. The dashed grey lines show the cosmological constant solution $w = -1$.

The most accurate current constraints on w_0 and w_a are [14],

$$w_0 = -1.04^{+0.72}_{-0.69} \quad (95\%;\ \text{Planck} + \text{WP} + \text{BAO})\,, \tag{1.56}$$

$$w_a < 1.32 \quad (95\%;\ \text{Planck} + \text{WP} + \text{BAO})\,. \tag{1.57}$$

Figure 1.2 shows the 2D marginalized likelihood contours for w_0 and w_a, for different combination of observational datasets. With the exception of the Planck+WP+SNLS dataset combination, where the case $w = -1$ lies on the 2σ boundary of the $w_0 - w_a$ distribution, the constraints are still consistent with a cosmological constant.

1.4 Formation of Structure in the Universe

One of the most challenging problems in modern cosmology is to understand the formation of structure in the Universe. From the observational stand point, the Universe seems to be statistically homogeneous and isotropic on large cosmological scales, but on the other hand, the existence of galaxies, clusters of galaxies, superclusters, voids and filaments is evidence of an irregular matter distribution on smaller scales. Within the standard model of cosmology, it is widely accepted that these complex structures observed today had their origin in the gravitational collapse of tiny primordial density inhomogeneities produced in the early Universe. The precise physical origin of these density inhomogeneities is still not fully understood, but, as we shall see in a later section (section 1.6), they may have resulted from quantum fluctuations generated during an inflationary epoch in the very early Universe [26]. The standard theory of structure formation studies small fluctuations around the FLRW metric so that the universe at large scales is still described as a homogeneous and isotropic background against which small inhomogeneities, representing the wide range of structures, stand out.

Jeans [70], in 1902, was the first to show that small density perturbations, $\delta\rho$, in a homogeneous and isotropic self-gravitating fluid, can grow over time. In fact, slightly overdense regions will exert higher gravitational effect in their surrounding, inducing the infall of material and as a consequence increasing their densities. As this occurs, pressure gradients may develop and resist the gravity forces, counteracting the collapse. The characteristic scale, where the two opposing forces cancel is known as the *Jeans length*, R_J. The processes, just described, became known as the gravitational Jeans instability and initially they were studied in an attempt to describe the formation of planets and stars in a static universe. However, it was not until 1946 by the hand of Lifshitz [71], that the theory of gravitational instability was applied to an expanding universe, thus presenting the first general analysis of the evolution of inhomogeneities in FLRW models using linear perturbation theory.

We can divide the growth of density perturbations in two regimes: the linear and the non-linear regimes. This distinction is done according to the value of the density contrast, written as,

$$\delta\left(\mathbf{x},t\right)\equiv\frac{\rho\left(\mathbf{x},t\right)-\rho_0\left(t\right)}{\rho_0\left(t\right)}, \tag{1.58}$$

which is a dimensionless quantity that quantifies, at each point in the spacetime, the deviation of the density field $\rho\left(\mathbf{x},t\right)$ relative to the mean density of the Universe, $\rho_0\left(t\right)=\left\langle\rho\left(\mathbf{x},t\right)\right\rangle$. As long as the perturbations are small, i.e. $|\delta|\ll 1$, first order perturbation theory can be used to describe the density perturbation. However, as perturbations grow larger, the accuracy of solutions found using this perturbative method starts to decrease and higher-order terms would need to be considered at later times. Eventually, even perturbative theory itself breaks down at the point when complex non-linear structures form and therefore other methods such as numerical techniques have to be employed to accurately follow the evolution of structure.

1.4.1 The linear perturbation theory of structure growth

The description of the evolution of the density perturbations in an expanding FLRW universe is usually a hard problem to tackle, since it requires the study of the behaviour of the perturbations for different components of the cosmic fluid on a large range of scales. To simplify the problem, it is assumed that perturbations are small during the initial stages of the process of structure formation. In this case, it is possible to write the equations governing their evolution in a perturbative form, neglecting all the non-linear terms. Furthermore, even though a detailed analysis of the growth of perturbations in a multi-component cosmological fluid requires a full relativistic approach (see [55]), if the perturbations are much smaller than the Hubble radius, the curvature of the Universe can be safely neglected and thus many fundamental aspects of the evolution of the perturbations can be followed by just using Newtonian mechanics. Having this in mind, the description of the evolution of the non-relativistic density field on cosmological scales can be done resorting to a set of three coupled differential equations, involving the gravitational potential ϕ, the peculiar velocity $\mathbf{v}$, the pressure p and the density contrast δ:

$$\nabla^2\phi = 4\pi G\rho, \tag{1.59}$$

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) \equiv \frac{D\rho}{Dt} + \rho\nabla\cdot\mathbf{v} = 0, \tag{1.60}$$

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} \equiv \frac{D\mathbf{v}}{Dt} = -\frac{1}{\rho}\nabla p - \nabla\phi, \tag{1.61}$$

where ρ, $\mathbf{v}$, p and ϕ are in general functions of the position $\mathbf{r}$ and t. Eqs. (1.59), (1.60) and (1.61) are respectively, the Poisson equation which relates the distribution of matter with the gravitational potential; the continuity equation which ensures mass conservation; and the Euler equation, describing the motion of the fluid element and ensures momentum conservation.

To proceed with the description of the perturbations, we expand the previous quantities in terms of their unperturbed (ρ_0, ϕ_0, p_0, $\mathbf{v}_0$) and perturbed values ($\delta\rho$, $\delta\phi$, δp, $\delta\mathbf{v}$), as:

$$\rho = \rho_0 + \delta\rho, \tag{1.62}$$

$$\mathbf{v} = \mathbf{v}_0 + \delta\mathbf{v}, \tag{1.63}$$

$$p = p_0 + \delta p, \tag{1.64}$$

$$\phi = \phi_0 + \delta\phi, \tag{1.65}$$

where $\delta\rho \equiv \rho_0\delta$ and the perturbation of the velocity field, $\delta\mathbf{v}$, is simply the peculiar velocity which measures deviations from the smooth Hubble expansion. The set of equations resulting from the substitution of eqs. (1.62 - 1.65) into eqs. (1.59 - 1.61), once written in Fourier space, are solved if one considers solutions in the form of plane waves:

$$f(\mathbf{r},t) = \int \frac{d^3k}{(2\pi)^3} f_k(t)\, e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{1.66}$$

where k and $f_k(t)$ are respectively, the wavenumber and the amplitude. Further assuming that the perturbations are adiabatic and dropping all non-linear terms, we obtain the following second-order differential equation describing the evolution of each Fourier mode of the density contrast:

$$\ddot{\delta_k} + 2H\dot{\delta_k}\left(\frac{c_S^2 k^2}{a^2} - 4\pi G\rho_0\right) = 0, \tag{1.67}$$

where $c_S^2 \equiv (\partial p/\partial\rho)_S$ is the adiabatic sound speed and $k = |\mathbf{k}|$ and dot represents the derivative with respect to time t. Eq. (1.67) allows to define the so-called *Jeans length* R_J and corresponding wavenumber as:

$$R_J = \frac{2\pi}{k_J}; \quad k_J = \frac{2a}{c_S}\sqrt{\pi G\rho_0}. \tag{1.68}$$

As mentioned before, the Jeans length is the characteristic scale where gravity forces and pressure gradients cancel each other, thus defining two different regimes for the evolution of the density contrast:

1. $R < R_J$ (or $k > k_J$): perturbations on these scales are not able to grow and form structures, behaving instead as acoustic waves.

2. $R > R_J$ (or $k < k_J$): this regime for the evolution of δ is much more interesting, since the perturbations on these scales are large enough to be gravitationally unstable, thus structures can form.

The homogeneous second-order differential equation (1.67), describing the evolution of the density perturbations in the linear regime, contains partial derivatives taken only with respect to time. In the regime where the condition $R > R_J$ is verified, it admits solutions of the form,

$$\delta(\mathbf{k}, t) = D_+(t)\,\delta_{0+}(\mathbf{k}) + D_-(t)\,\delta_{0-}(\mathbf{k}), \tag{1.69}$$

where $\delta_{0\pm}(\mathbf{k})$ is the density perturbation at a specific time (e.g. today). The functions D_+ and D_- are the growing and decaying modes respectively. In the case of an Einstein-de Sitter universe, i.e. $\Omega_{m0} = 1$ and $k = 0$, (1.67), D_+ and D_- are given by,

$$D_+ \propto t^{2/3} \propto a, \tag{1.70}$$

$$D_- \propto t^{-1} \propto a^{-3/2}. \tag{1.71}$$

Since, D_- decreases over time, the growing mode D_+ also known as *linear growth factor*, is the relevant quantity for structure formation and thus eq. (1.69) can be further simplified to,

$$\delta(\mathbf{k}, t) = D_+(t)\,\delta_{0+}(\mathbf{k}). \tag{1.72}$$

In general the linear growth factor D_+ has to be obtained numerically, however for a flat ΛCDM universe, it can be written as (see e.g. [72]),

$$D_+ \propto \frac{H(t)}{H_0} \int_0^a \frac{dx}{\left[\Omega_{m0}/x + \Omega_{\Lambda 0}/x^2 + (1 - \Omega_{m0} - \Omega_{\Lambda 0})\right]^{3/2}}, \tag{1.73}$$

with an additional constraint such that $D_+(t_0) = D_+(a = 1) = 1$, which implies that $\delta_{0+}(\mathbf{k})$ is today's density contrast according to linear perturbation theory.

1.4.2 Statistical properties of the density field

Within the framework of cosmic structure formation, it is not possible to predict the precise value of the density field $\delta(\mathbf{x})$ at each point in space. Instead we are only able to estimate its statistical properties as a function of time. In cosmology, fields such as the matter density are assumed to be random variables, whose properties are completely described through its moments, $\langle \delta(\mathbf{x}) \rangle$, $\langle \delta^2(\mathbf{x}) \rangle$, ..., $\langle \delta^n(\mathbf{x}) \rangle$. These are also known, respectively, as the one-point, two-point, ..., n-point correlation functions of the density field, which can be computed as,

$$\langle \delta^n(\mathbf{x}) \rangle = \int d\delta(\mathbf{x})\, \delta^n(\mathbf{x})\, p\left[\delta(\mathbf{x})\right], \tag{1.74}$$

where the brackets denote averages over an ensemble of possible universes and $p\left[\delta(\mathbf{x})\right]$ is the probability density function (hereafter PDF). Since we are limited to our own Universe, these distributions are inaccessible to us, at least in principle. In practice, it is always assumed that cosmological random fields are ergodic and therefore the true PDF can be estimated by replacing the ensemble averages by spatial averages over very large volumes of our Universe. Another common assumption is that cosmological random fields are invariant under rotations and translations, in the statistical sense. This implies, for example, that the one-point PDF is independent of position $\mathbf{x}$ and the two-point correlation function, that correlates two distinct points in space $\mathbf{x}$ and $\mathbf{x}'$, is only a function of the distance $r = |\mathbf{x} - \mathbf{x}'|$. An immediate consequence of ergodicity is that the one-point function of the density field is zero, i.e. $\langle \delta(\mathbf{x}) \rangle = 0$. The two-point correlation function can be defined as

$$\langle \delta(\mathbf{x})\, \delta(\mathbf{x}') \rangle = \xi\left(|\mathbf{x} - \mathbf{x}'|\right) = \xi(r), \tag{1.75}$$

with $\mathbf{x}' = \mathbf{x} + \mathbf{r}$ and we have made the assumption that the density field is homogeneous and isotropic. Fourier transforming the density field $\delta(\mathbf{x})$ and using eq. (1.75), we can write,

$$\langle \delta(\mathbf{k})\, \delta(\mathbf{k}') \rangle = (2\pi)^3\, \delta_D\left(\mathbf{k} + \mathbf{k}'\right) P_\delta(k), \tag{1.76}$$

where δ_D is the 3D Dirac function and P_δ is the power spectrum of the overdensity field, defined as the Fourier transform of the two-point correlation function,

$$P_\delta(k) = \int d^3r\, \xi(\mathbf{r})\, e^{-i\mathbf{k}\cdot\mathbf{r}}, \tag{1.77}$$

The power spectrum is usually redefined in a dimensionless form, as the variance per (natural) logarithm interval,

$$\mathcal{P}_\delta \equiv \frac{4\pi k^3}{(2\pi)^3} P_\delta, \tag{1.78}$$

and using the above redefinition, the variance of the overdensity field (the two-point correlation function for $r = 0$) is

$$\sigma_\delta^2 \equiv \langle \delta^2 (\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} \mathcal{P}_\delta. \tag{1.79}$$

Most models of inflation favour density perturbations of adiabatic nature, with nearly Gaussian probability distribution with variance σ_δ^2 and a primordial density power spectrum $\mathcal{P}$ well approximated by a power law,

$$\mathcal{P}_\delta = A k^{n_s - 1}, \tag{1.80}$$

where n_s is the so-called spectral index of scalar perturbations and A is a normalization constant, to be inferred from observations. In the particular case where $n_s = 1$, one obtains a scale-invariant power spectrum, also known as Harrison–Zel'dovich power spectrum [73, 74], and thus the fluctuations in the density field are scale-independent. Current observational constraints both on normalization constant and the spectral index of the primordial power spectrum are, respectively, $n_s = 0.9616 \pm 0.0094$ and $\ln\left(10^{10} A\right) = 3.103 \pm 0.072$ [14].

Another quantity useful to describe the statistical properties of the density field, which will be very handy later on in this thesis, is the smoothed density contrast field, $\delta_R (\mathbf{x})$, defined as the convolution of the density field with some window function $W (R, r = |\mathbf{x} - \mathbf{x}'|)$ (the smoothing kernel), which tends to zero quickly for $r > R$,

$$\delta_R (\mathbf{x}) = \int d^3 x' W \left(R, r = |\mathbf{x} - \mathbf{x}'|\right) \delta (\mathbf{x}) . \tag{1.81}$$

Smoothing the density is important, since many observational procedures automatically introduce smoothing. In this thesis we will make use of the top-hat filter, whose Fourier transform, $W_R (k)$ is given by,

$$W_R (k) = 3 \left[\frac{\sin (kR)}{(kR)^3} - \frac{\cos (kR)}{(kR)^2} \right]. \tag{1.82}$$

The smoothing scale R is often specified by the mass within the volume defined by the window function at the present time. In the case of the top-hat filter this is $M = 4\pi \rho_{m0} R^3 / 3$. It is also possible to compute the variance of the density field smoothed by the top-hat filter,

$$\sigma_R^2 \equiv \langle \delta_R^2 (\mathbf{x}) \rangle = \int_0^\infty \frac{dk}{k} W_R^2 (k) \mathcal{P}_\delta. \tag{1.83}$$

Using the previous expression, we can estimate the normalization of the power spectrum, which is often expressed in terms of the variance of the density inside spheres of radius $8h^{-1}$Mpc, commonly known as σ_8.

Before concluding this subsection and for reasons that will be clear later on, it is important to define the three-point correlation function for the density field, as

$$\Upsilon = \langle \delta(\mathbf{x})\, \delta(\mathbf{x}+\mathbf{r})\, \delta(\mathbf{x}+\mathbf{s}) \rangle, \tag{1.84}$$

where the ensemble average is taken over all possible triangle configurations with sides $r, s, t = |\mathbf{r} - \mathbf{s}|$. Furthermore, we can write the analogous of eq. (1.76),

$$\langle \delta(\mathbf{k})\, \delta(\mathbf{k}')\, \delta(\mathbf{k}'') \rangle = (2\pi)^3\, \delta_D(\mathbf{k}+\mathbf{k}'+\mathbf{k}'')\, B_\delta(k, k'k''), \tag{1.85}$$

with $B_\delta(k, k'k'')$ being the bispectrum of the density field. This quantity plays a significant role in constraining primordial non-Gaussianities, since it is the lowest order statistics sensitive to deviations from Gaussian initial conditions which may be present in the primordial matter density fluctuations.

1.4.3 The non-linear evolution of density perturbations

The linear perturbation theory, discussed earlier, is valid if perturbations are small, i.e, $|\delta| \ll 1$, condition which is roughly guaranteed everywhere in space as long as the rms fluctuations of the density contrast are much smaller than the unity, $\sigma(t) \ll 1$. On the other hand, regions where $\delta(\mathbf{x}, t) \gtrsim 1$ are very rare initially, but as more and more perturbations collapse, the linear perturbation theory becomes inappropriate to describe the ever increasing number of such regions. A way around the later limitation to extend the use of linear perturbation theory is to apply the linear evolution equations to the density contrast smoothed on a scale R, i.e. $\delta_R(\mathbf{x}, t)$, instead of $\delta(\mathbf{x}, t)$. Again, this is valid as long as $\sigma(R, t) \ll 1$, in which case the wavenumbers k smaller than $1/R$ evolve linearly. Obviously, this is only handy if the scales we are interested in are larger than the smoothing scale R.

A logical way to describe the evolution of the perturbations in their initial stages of non-linear evolution (quasi-linear regime) is to resort to second order terms in the evolution equations of δ [61, 75]. However, this method has also limited applicability, since the amplitude of even higher order terms become important quite rapidly. The transition from the quasi-linear regime to the non-linear evolution stages of density perturbations is in general considered to occur when $\sigma(R, t) \backsim 1$.

Numerical N-body simulation techniques are a good way to track the full evolution of perturbations. However, there are few analytical approximations, that have the advantage to be generally more flexible and a lot less time consuming than numerical simulations. Such analytical approximations still allow us to get valuable insights on the general problem of describing perturbations during the first stages of their non-linear evolution. Some of these methods will be reviewed next.

1.4.4 The spherical collapse model

We can use the spherical collapse model to estimate how dense and when a spherically symmetric overdense distribution of matter abandons the general expansion of the universe and starts to

gravitationally collapse towards its centre (see for example [61, 75, 76]) .

When the density of an overdense region becomes slightly larger than the critical density it starts to behave as a small closed universe, which evolves almost independently of the outside space. At the beginning, the sphere will continue to expand until it reaches a maximum radius, which is known as the turnaround point, and finally collapses after a time interval, t_{coll}, equal to double the time it takes to get to maximum expansion, t_{max}. In an Einstein–de Sitter universe, it can be shown that inside the sphere, the linear evolution of the density perturbation is given by $\delta_{lin} = 3/20\,(6\pi t/t_{max})^{2/3}$, which implies that the linear density at turnaround and collapse are given, respectively, by:

$$\delta_{lin}^{turn} = 3/20\,(6\pi)^{2/3} \simeq 1.06, \tag{1.86}$$

$$\delta_{lin}^{coll} = 3/20\,(12\pi)^{2/3} \simeq 1.686. \tag{1.87}$$

It is also possible to compute the non-linear density at both these instants. At turnarround, the model predicts

$$1 + \delta_{nonlin}^{turn} = \left(\frac{3\pi}{4}\right)^2 \backsimeq 5.55, \tag{1.88}$$

while for the time of collapse, the spherical model predicts an infinite density. Note however, that these results do not take into account the pressure forces resisting the gravitational infall. Yet, in reality, one should expected that part of the kinetic energy of collapse will be transformed into random motion of particles and eventually the object will reach a state of virial equilibrium characterized by a temperature, T_{v}, and radius, R_{v}. Resorting to the viral theorem to compare the virial energies at the time of maximum expansion and compression, we find that $R_{\mathsf{v}} = R_{max}/2$, where R_{max} is the radius of the spherical region at maximum expansion. High precision numerical simulations have shown that R_{v} is indeed an accurate estimate of the radius of the collapsed sphere and that δ_{lin}^{coll} is a good estimative of its density when it becomes virialized. Assuming that the object virializes with the radius R_{v} at the time of collapse $t_{coll} = 2t_{max}$, then the non-linear overdensity is given by

$$1 + \delta_{nonlin}^{vir} = 18\pi^2 \simeq 178, \tag{1.89}$$

consistent with numeric N-body simulations results (see e.g. [77]).

The predictions of the spherical collapse model change slightly for non Einstein-de Sitter universes. Both the linear and non-linear overdensities at collapse are in general a function of the matter density, Ω_{m0}, but for low-density models, i.e. $\Omega_{m0} \sim 0.1 - 1$, the variations in δ_{lin}^{coll} are less than $\sim 5\%$. Moreover, the overdensity at collapse must be corrected by $\Omega_{m0}^{-0.7}$ for $\Omega_{\Lambda 0} = 0$ universes and $\Omega_{m0}^{-0.45}$ for flat universes with a non-vanishing cosmological constant Λ (see early works [78–80]). Often, the non-linear overdensity is expressed in terms of the critical density instead of the background density, $\Delta \equiv \rho\,(t)\,/\rho_{crit} = (1 + \delta_{nonlin})\,\Omega_m$, and it is usually referred to as the non-linear density contrast. For detailed calculations involving the evolution of the density contrast at collapse, $\Delta_c\,(z)$, there is the following fitting formula, from [81]:

$$\Delta_c(z) = \begin{cases} 18\pi^2 + 60x - 32x^2, & \Omega_{\Lambda 0} = 0 \\ 18\pi^2 + 82x - 39x^2, & \Omega_{m0} = 1 - \Omega_{\Lambda 0} \end{cases} \tag{1.90}$$

with $x = \Omega_m(z) - 1$. In the range of values, $\Omega_m \sim 0.1 - 1$, expression (1.90) is accurate to 1%.

1.4.5 Zel'dovich approximation

The Zel'dovich approximation, [74, 82], is a very successful first–order Lagrangian perturbation theory, which describes perturbations with respect to particle positions rather than to the density field (as described in the Eulerian perturbative method). In this framework, particles move in straight lines after being displaced from their original (comoving) positions by the action of an initial velocity field sourced by the density perturbations. Hence, the comoving position and peculiar velocity of a given particle at the time t are:

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{q} + D_+(t)\,\mathbf{u}(\mathbf{q}) \\ \dot{\mathbf{x}}(t) &= \dot{D}_+(t)\,\mathbf{u}(\mathbf{q}) \end{aligned} \tag{1.91}$$

In the "jargon" of fluid mechanics, $\mathbf{x}$ and $\mathbf{q}$ are said to be the Eulerian and Lagrangian coordinates, respectively. In addition, in previous expressions, it common to assume that $\mathbf{u}(\mathbf{q})$ is an irrotational field, satisfying the conditions,

$$\begin{aligned} \delta(\mathbf{q}) &= -\nabla \cdot \mathbf{u}(\mathbf{q}) \\ \nabla \times \mathbf{u} &= 0 \end{aligned} \tag{1.92}$$

The Zel'dovich approximation is commonly used to set up the initial conditions in numerical N–body simulations and it will be useful later on in this thesis. The initial velocity field is computed from eqs. (1.92) by generating a random Gaussian realization of the density power spectrum. Afterwards, eqs. (1.91) are used to generate the appropriate initial matter and velocity distributions, required for the first integration step of the simulations.

Despite its usefulness, the Zel'dovich approximation has some limitations. The trajectories of the particles are uniquely set by the initial displacements and, as motions intercept, particles keep moving away from the forming structures. At this point, usually referred to as shell crossing, the Zel'dovich approximation breaks down. However, being a first-order Lagrangian perturbative method, the Zel'dovich approximation usually achieves an acceptable level of accuracy in describing the perturbations for longer than in the (Eulerian) linear perturbative case.

1.4.6 The Press-Schechter Theory

The Press-Schechter formalism (hereafter PS formalism) [83] is perhaps one of the most useful and outstanding analytical methods used to describe the density perturbations in their early stages

of non-linear evolution. Eventually, density perturbations will, under the influence of gravitational instability, collapse and form virialized structures such as dark matter halos. The estimation of the abundance of these objects as a function of redshift and mass depends for instance, on the initial power spectrum, the expansion rate of the Universe, astrophysical processes such as dissipation and feedback. Furthermore, a robust observational proxy for the mass is also indispensable.

Assuming that the initial density perturbation field follows a Gaussian statistics, the PDF of the density filtered on a comoving scale R (associated to a mass scale M), $\delta_M \equiv \delta(\mathbf{x}, R)$ is given by,

$$P(\delta_M)\, d\delta_M = \frac{1}{\left(2\pi\sigma_M^2\right)^{1/2}} \exp\left(-\frac{\delta_M^2}{2\sigma_M^2}\right) d\delta_M, \tag{1.93}$$

with σ_M^2 being the variance of the smoothed density contrast filtered on the comoving scale R defined previously in section 1.4.2. The probability of a density fluctuation exceeding a given threshold δ_c is,

$$P_{>\delta_c}(M) = \int_{\delta_c}^{\infty} P(\delta_M)\, d\delta_M. \tag{1.94}$$

The PS formalism assumes that the number density of virialized halos formed is proportional to the probability given by eq. (1.94), thus characterizing density perturbations greater than δ_c. From the spherical collapse model, discussed in subsection 1.4.4, a halo forms when the density contrast reaches the value $\delta_c = 1.686$ (where we have set $\delta_c \equiv \delta_{lin}^{coll}$, see eq. (1.87)). The number of these objects, at a given redshift z, within the mass interval $[M, M + dM]$, will therefore be proportional to the difference between $P_{>\delta_c}(M)$ and $P_{>\delta_c}(M + dM)$. This quantity is known as the *mass function*. An analytical expression of the mass function, $n(M)$, was derived by [83], by means of the ansatz,

$$n(M)\, MdM = 2\overline{\rho}_m \left[P_{>\delta_c}(M) - P_{>\delta_c}(M + dM)\right], \tag{1.95}$$

where the inclusion of the factor 2 is to account for the so-called *cloud-in-cloud problem* (see e.g. [84, 85]). Using eqs. (1.94) and (1.95), one can write a mathematical expression for the mass function in terms of the PDF (1.94),

$$\frac{dn}{dM}(z, M) = -2\frac{\overline{\rho}_m}{M}\frac{d}{dM}\left[\int_{\delta_c(z)/\sigma_M}^{\infty} d\nu P(\nu)\right], \tag{1.96}$$

where $\overline{\rho}_m$ is the average comoving matter density and $\nu \equiv \delta_M/\sigma_M$. The redshift dependence of the mass function has been incorporated in the threshold for collapse $\delta_c(z) = 1.686/D_+(z)$ (note that $D_+(z)$ is the linear growth factor normalized to unity at $z = 0$, defined in section 1.4.1). Inserting the PDF in eq. (1.93) into the expression given by eq. (1.96), we then recover the well known result obtained in [83],

$$\frac{dn_{PS}}{dM}(z, M) = -\sqrt{\frac{2}{\pi}}\frac{\overline{\rho}_m}{M^2}\frac{\delta_c(z)}{\sigma_M}\frac{d\ln\sigma_M}{d\ln M}\exp\left[-\delta_c^2(z)/\left(2\sigma_M^2\right)\right]. \tag{1.97}$$

The Press-Schechter mass function has been tested against numerical N-body simulations and results show that the PS formalism under-predicts the abundance of high mass halos and over-predicts the low-mass ones, which is not surprising, given that the spherical collapse assumption invoked to obtain eq. (1.97), may not hold in detail [86]. Thus, in order to close the gap between the predictions of the PS formalism and simulations, Sheth and Tormen [86] suggested the fitting formula for the mass function, assuming an ellipsoidal collapse,

$$\frac{dn_{ST}}{dM}(z,M) = -\sqrt{\frac{2a}{\pi}}A\left(1+\left(\frac{a\delta_c^2(z)}{\sigma_M^2}\right)^{-p}\right)\frac{\overline{\rho}_m}{M^2}\frac{\delta_c(z)}{\sigma_M}\frac{d\ln\sigma_M}{d\ln M}\exp\left[-\delta_c^2(z)/\left(2\sigma_M^2\right)\right], \tag{1.98}$$

with $a = 0.707$,$A = 0.3222$ and $p = 0.3$. The Sheth and Tormen mass function given in eq. (1.98), reduces to the original Press-Schechter form for $a = 1$, $A = 0.5$ and $p = 0$.

At this point, it makes sense to define an expression to estimate the total number of virialized halos, per redshift interval, above some mass threshold, M_{lim},

$$\frac{dN}{dz}(z, M > M_{min}) = f_{sky}\frac{dV(z)}{dz}\int_{M_{lim}}^{\infty} dM\frac{dn}{dM}, \tag{1.99}$$

where f_{sky} if the fraction on the sky being observed and the $dV(z)/dz$ is the comoving volume element given by,

$$\frac{dV}{dz}(z) = 4\pi\,\chi(z)\,\frac{d\chi}{dz}(z), \tag{1.100}$$

with $\chi(z)$ being the comoving radial distance as a function of redshift.

1.4.7 Numerical Methods

Although it is possible to get a reasonable physical understanding of the processes involved in structure formation from analytical techniques, they rely on approximations, thus providing limited solutions only suited for special symmetries and configurations. Moreover, the study of high order perturbation theory can become a very difficult problem to tackle and thus analytical models may not yield detailed and accurate theoretical predictions that can be tested against observations. Therefore, in order to follow the full non-linear evolution of structure growth, one must rely on more general methods such as numerical simulations.

Nowadays, simulations have become an indispensable tool to construct and describe a standard model of formation and evolution of the structures in the Universe: the so-called ΛCDM model, which represents a major theoretical paradigm of modern cosmology. Remarkable advances in the study of the temperature profile of galaxy clusters, the non-linear parts of the clustering of dark matter, the density profile of virialized dark matter halos, just to name a few, benefited greatly from numerical simulations. The development of highly sophisticated numerical algorithms combined with the continuous growth of computer power and performance, are fundamental requirements to achieve accurate theoretical predictions with simulations. In the following, we will provide a short summary of some of the basic ideas of numerical simulations in the context of cosmological structure formation.

Due to the fact that a large fraction of the mass in the Universe is made of dark matter, an accurate modelling of this collisionless component is essential. The N-body technique is employed for this purpose. According to this methodology, a portion of the Universe is represented by means of a large number N of particles, moving under the influence of their mutual gravitational interaction. N-body simulations require the accurate tracking of the trajectories of a set of particles interacting within a rectangular box (most often cube), assumed to be representative of the Universe as a whole. This is translated in the discretization of the mass resorting to N point-like particles and performing the integration of their equations of motion from some instant in the past till the present time.

The detailed procedure is essentially standard, in the sense that it identically applies to whatever cosmological model one wishes to simulate: initial conditions (positions and velocities) of the distribution of particles encode the information on a specific cosmology, while the N-body integrator acts on these initial conditions at the beginning of the simulation and evolves them forwards in time. However, setting-up the initial conditions poses a challenge on its own, since inaccuracies or numerical artefacts may survive and be imprinted on the non-linear evolution of structures till the end of the computation.

The algorithm used to compute the gravitational force has a strong impact on the efficiency of the N-body simulations. As a first approach, one can compute the force acting on a single particle by simply summing up the contribution of all other the particles in the simulations. This is known as the Particle-Particle (PP) algorithm and the Newton force equation can be written as,

$$\mathbf{F}_j = \sum_{i \neq j}^{N} \frac{G m_i m_j}{\left(\mathbf{r}_{ij}^2 + \epsilon_s^2\right)^2}, \tag{1.101}$$

where a softening length, ϵ_s, is introduced in order to suppress two-body forces at very small separations. The parameter is required to be present to make sure that two-body relaxation is not important and also that no bound two-particle pairs can form. Moreover, ϵ_s can also be regarded as a spatial resolution limit, i.e, structures smaller than this scale cannot be represented.

The PP-method is very accurate, however it is not able to handle large systems with a number of particles greater than $\sim 10^5 - 10^6$, since the computational time scales as $O\left(N^2\right)$. The Particle-Mesh (PM) method offers a solution to this problem. The PM method works by computing the gravitational potential on a Cartesian grid with periodic boundary conditions, taking advantage of the Fast Fourier Transform (FFT) algorithms to speed up the solution of the Poisson equation in Fourier space. By doing this, the number of operations scales as $N + N_g \log N_g$, where N and N_g are, respectively, the number of particles and the number of grid points. However, at smaller scales, the PM method has a poor force resolution, being limited by the finite size of the grid; which is translated in a serious limitation for the strong clustering occurring during cosmic structure formation. In 1981, Hockney [87] merged the accuracy of the PP method with the computational efficiency of PM, in a novel scheme known as the P^3M approach: where the long range forces are computed on a grid, while the contribution due to the nearby particles, which change rapidly with time, is treated with the PP method.

N-body numerical simulations underwent another efficiency revolution resulting from the adoption of Hierarchical Tree algorithm [88, 89]. In the Tree methods scheme, space is divided recursively into a hierarchy of cells, each containing one or more particles. The structure inside a cell of size l and distance d (from the point where forces need to be computed) can be safely discarded as long as the condition $d > l/\theta$ (where θ is some accuracy parameter) holds, and therefore its gravitational field is well approximated with a multipole expansion. On the other hand, if the former conditions do not hold, space is further divided in sub cells, if necessary up the point where the smallest cells contain at most one particle. These previous operations reduce the total computation time of the scheme, since the summation over all particles is replaced with a much shorter set of low-order multipole expansions of the gravitational potential of particle groups.

Up until now, only the numerical evolution of the collisionless material in the Universe have been addressed. In order to model the visible matter, one must study the joint dynamics of the baryonic and non-baryonic components and track the hydrodynamical processes besides gravity. Indeed, both radiative and hydrodynamical effects cannot be neglected, since they have a deep impact on the galaxy formation process. The simplest case of non-relativistic, non-radiative simulations of an ideal gas is a reasonable approximation for the hot plasma in clusters of galaxies. If more realism in the simulations is needed, one can go one step further and incorporate radiative cooling and heating processes, as well as feedback to prevent excessive gas cooling.

The Smoothed Particle Hydrodynamic (SPH) technique is a powerful approach used to model hydrodynamical processes. In this scheme, the basic fluid equations are solved numerically by means of performing an ensemble of discrete particles, characterized by their positions, velocities and internal thermodynamic variables (like the thermal energy per unit mass). A spatial smoothing length L is associated with the particles, over which their properties are interpolated by a kernel function. Since SPH is a Lagrange method, it allows to accurately explore and model the self-gravity together with hydrodynamics, thus being particularly well suited for cosmological simulations of structure formation, where an automatically adaptive resolution is required due to the high collapse factors of the mass.

Later on in this thesis, the simulation techniques briefly described above will be used, all performed using the Lagrangian code GADGET-2 [90, 91]. In this simulation code, the gravitational interaction is computed with a Tree method, combined with a PM approach for long-range forces, yielding the so-called TreePM method. The hydrodynamical processes are modelled with the SPH technique.

1.5 Problems of the Standard Cosmological Model

In previous sections, the Standard Model of Cosmology was extensively described. This remarkable framework allows us to describe with tremendous precision the physical processes that gave rise to the Universe we observe today and to its complex nature. Despite this amazing achievement, modern cosmology still faces some outstanding problems. In the next paragraphs, we will briefly highlight a few of these unsolved problems and motivate the Inflationary paradigm as a

possible solution to some of them.

The Horizon problem The Cosmological Principle and the existence of a particle horizon in the Friedmann models seem to be fundamentally incompatible, unless we postulate that large-scale homogeneity and isotropy of the Universe is part of its initial conditions. The most remarkable observational evidence in favour of the Cosmological Principle is the uniformity of the temperature field of the cosmic radiation background. This can only be explained if, at the time the radiation last scattered, all regions of the Universe were in causal contact, which would smooth out any temperature fluctuation.

Since the distance travelled by a light signal between the Big Bang and the present is finite, then a set of points are said to be causally connected with an observer "O", if at some instant t the observer is able to receive light signals emitted from those points. For this to occur, the points must be located inside a sphere centered upon O, with proper radius:

$$R_H(t) = a(t) \int_0^t \frac{dt'}{a(t')}. \tag{1.102}$$

If the integral on the right-hand side of eq. (1.102) converges to a finite value at $t = 0$, then the quantity R_H exists and it is called the "particle horizon" at time t. In conventional Friedman models dominated at early times by radiation, such horizon exists. On the other hand, the distance travelled by a photon emitted at the epoch of last scattering, i.e. $z_{ls} \sim 1000$, and detected by an observer located at t_0 is related to the proper comoving distance by,

$$r_{ls} \approx \frac{a(t_{ls})}{a(t_0)}\chi = \frac{t_0 - t_{ls}}{1 + z_{ls}} \approx \frac{t_0}{z_{ls}}. \tag{1.103}$$

However, at this epoch, the particle horizon radius is given instead by,

$$R_H(t_{ls}) \approx 3t_0 z_{ls}^{-3/2} \approx 10^{-3/2} r_{ls}, \tag{1.104}$$

and thus $R_H < r_{ls}$. This means that an observer located at t_0 sees different regions on the last scattering surface that are causally disconnected (since they have non-overlapping horizons), which is in tension with observations.

The flatness problem The framework of standard cosmology is not able to explain the so-called flatness problem. As seen before, performing some mathematical manipulations to the Friedman equation eq. (1.10), we can write the total energy density of the Universe at a given time t as,

$$\Omega(t) - 1 = \frac{K}{(aH)^2} \tag{1.105}$$

In a matter and radiation dominated universe, the function aH decreases over time and therefore Ω deviates from unity quite quickly. If we divide, (1.105) by itself at present times it is easy to realize that a high degree of fine tuning is needed in the past so that $\Omega \backsim 1$. Around the Planck time, i.e. $t_p \approx 10^{-43}$ a time scale well beyond the applicability of classical GR equations, Ω can only deviate

from unity less than one part in 10^{60}. Clearly, $\Omega = 1$ is an unstable situation for the universe and any possible deviations larger than the one allowed by eq. (1.105) inevitably leads to a universe quite different from the one we observe today. So the question arises: why was the initial energy density of the universe so close to one?

Structure Formation and Small-scale inhomogeneity Despite the tremendous observational support for homogeneity and isotropy of the Universe at large scales, locally, the distribution of matter is quite anisotropic. In fact on smaller scales, the Universe in populated with galaxies, galaxy clusters, voids and filaments, which are thought to originate from tiny perturbations in the primordial density field that collapsed due to self gravitational instability. Although such tiny initial inhomogeneities are needed to justify the formation of structure that has taken place throughout the Universe, their existence and the physical process responsible for producing them is not accounted for in the Standard Cosmological Model. Instead, the primordial perturbations are postulated already showing the correct amplitudes on all scales, so that gravitational instability can correctly reproduce the present-day structures.

Missing relics At early times, the Universe had the ideal conditions so that a variety of stable, super-heavy particles species, predicted by particle physics, could be produced. Such particles, such as magnetic monopoles would have been able to survive throughout the evolution of the Universe till today in quite large quantities (the expected number density is roughly 10^{60}). However, observational evidence for this particles has not been found yet. This could either be due to wrong predictions made by particle physics or, somehow, there was a physical mechanism in the early Universe that diluted their density and thus, a piece of the Big Bang puzzle may be missing.

1.6 The Inflationary paradigm

The Inflationary Paradigm was introduced by Alan Guth in the 80's [26] not as a replacement to the Standard Cosmological Model, but as a complement, in order to solve its open problems. Inflation refers to a brief period of accelerated expansion of the Universe before the radiation era. Therefore, it only describes the initial stages of the Universe, being the posterior evolution described by the "traditional" Standard Cosmological Model.

It is commonly accepted that Inflation is driven by the energy of a scalar field ϕ called the inflaton, which has negative pressure and whose potential represents the energy density of the vacuum. During the epoch of inflation, the scalar field ϕ experiences a phase transition, rolling down its self-potential, from its highly unstable false vacuum state to the more stable state of true vacuum. In order for Inflation to solve the problems of cosmological model, it has to last for some time and thus the inflaton's potential must be sufficiently flat. After the inflaton field reaches the minimum of the potential, it starts to oscillate and decays, giving rise to a "hot soup" of baryonic particles and photons, heating the universe (reheating) in the process.

Although the inflationary paradigm was initially suggested as a possible explanation to the fact that there is no observational evidence for the existence of the unwanted relics predicted by the Grand Unification Theory (GUT) models, its most useful feature is that it provides a natural framework for the production of the primordial seeds, indispensable to give rise to the observable structures.

Dynamics of inflation The simplest model of inflation can be driven by a single scalar field, $\phi(\mathbf{x}, t)$, the inflaton, coupled with gravity through the metric $g_{\mu\nu}$ and whose action is given by

$$S_\phi = \int d^4x\sqrt{-g}\left(\frac{M_P^2}{2}R - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi)\right) \tag{1.106}$$

where $M_P = (8\pi G)^{-1/2}$ is the Planck mass, R is the Ricci scalar and $V(\phi)$ is the inflation potential. An insight on the dynamics of the inflaton is obtained from its equations of motion, which can be obtained by performing a variation of the action (1.106) with respect to the scalar field. Further assuming a FLRW metric (1.1), one arrives to the following equation:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\nabla^2\phi}{a^2(t)} + V_{,\phi} = 0 \tag{1.107}$$

where $V_{,\phi} = dV/d\phi$ and H assumes the form

$$H^2 = \frac{1}{3M_P^2}\left[V(\phi) + \frac{1}{2}\dot{\phi}^2\right] \tag{1.108}$$

Eq. (1.107) can be interpreted as describing the rolling down of the field ϕ towards the minimum of its potential at ϕ_0, under the effect of the force given by $dV/d\phi$ and against a source of friction given by the term H.

In the particular case of a homogeneous inflaton field, i.e., $\phi(\mathbf{x}, t) \equiv \phi(t) \Rightarrow \nabla^2\phi = 0$, behaving like a perfect fluid, it is possible to derive the equation for the density and pressure of the scalar field similar to the ones of eq. (1.39), from the zero and i-th order of the energy-momentum tensor. In turn, eq. (1.107) simplifies to,

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0. \tag{1.109}$$

The full steps leading to the derivation of the above equations, can be easily found in most cosmological textbooks and reviews, such as [55, 92, 93].

Inflation can be studied by means of the *slow-roll* approximation, which discards the first and last terms of eqs. (1.109) and (1.108), respectively, which yields to,

$$3H\dot{\phi} \simeq -V_{,\phi} \tag{1.110}$$

$$H^2 \simeq \frac{V(\phi)}{3M_p^2} \tag{1.111}$$

To be valid, previous approximations require that

$$\epsilon(\phi),\ |\eta(\phi)| \ll 1 \tag{1.112}$$

where ϵ and η are the known as the slow-roll parameters, which can be written in term of the inflationary potential and respective derivatives,

$$\epsilon(\phi) \equiv \frac{M_p^2}{2}\left(\frac{V_{,\phi}}{V}\right)^2 \tag{1.113}$$

$$\eta(\phi) \equiv M_p^2\left(\frac{V_{,\phi\phi}}{V}\right) \tag{1.114}$$

The small size of the slow-roll parameters, ϵ and η, constrain the potential $V(\phi)$ to be nearly flat, which is a necessary condition for the slow-roll approximation to be valid. However, the flatness of the potential is not a sufficient condition, if no prerequisite is done on the size of the kinetic term in $\dot{\phi}$, which in principle can be as large as required. Therefore, another condition is that the solution for a given potential must verify eq. (1.110).

Whenever conditions given by eq. (1.112) hold and thus $\ddot{\phi}, \dot{\phi} \simeq 0$, the inflaton field goes through a period of slow-roll, where pressure and density are related to each other by $p_\phi = -\rho_\phi \simeq V(\phi) =$ const.. Furthermore, since initially $\rho \propto a^{-4}$ while ρ_ϕ remains constant, at some time t_i after the Big Bang, the energy density of the inflaton field dominates the energy content of the Universe and thus drives the dynamics of the expansion. This corresponds to the start of the inflationary period. The scale factor in this era is given by,

$$\frac{a(t)}{a_i} = \begin{cases} \cosh\left[H(t-t_i)\right] & \Omega > 1 \\ \exp\left[H(t-t_i)\right] & \Omega = 1 \\ \sinh\left[H(t-t_i)\right] & \Omega < 1 \end{cases} \tag{1.115}$$

with $a_i = a(t_i) = (3/8\pi GV)^{1/2}$ and the Hubble expansion rate given by $H = (8\pi GV)/3 \propto$ const. , exhibiting a similar dynamical behaviour as the de Sitter Universe. Soon after some time interval $t - t_i \gg (3/8\pi GV)^{1/2}$, regardless of the geometry, the scale factor starts to grow exponentially and the total energy density becomes very close to unity, ($\Omega = 1$), thus becoming flat. Hence, the undesirable fine-tuning condition on Ω required by the standard Hot Big Bang scenario can be safely solved. But the inflationary period cannot last forever. Eventually, the universe needs to re-enter in a decelerating phase, in order for the observed light elements abundances to form, in a process known as primordial nucleosynthesis. Otherwise, the accelerated expansion would lead to an empty Universe, without any inhabitants to contemplate its emptiness.

The inflationary process ends with the scalar field reaching the minimum of its potential. At this point, the inflaton ends its slow-roll motion and it starts to oscillate around the minimum of the potential. Moreover, as the inflaton field oscillates, the particles associated with it decay and the resulting energy is transferred to the cosmic fluid. Thus, the later experiences a fast temperature increase, in a process known as reheating. After the reheating phase has taken place, the Universe

enters in a radiation dominated epoch and carries on with its standard evolution. Typically, inflation occurs just before the GUT phase transition approximately at $t_i \backsim 10^{-34}$s and ends around $t_i \backsim 10^{-12}$s. It is possible to compute the amount of inflation produced during this short period of time, which is expressed in terms of the number of *e-foldings*,

$$\mathcal{N} = \ln\left(\frac{a\left(t_f\right)}{a\left(t_i\right)}\right) \simeq \int_{t_i}^{t_f} H dt \simeq -8\pi G \int_{\phi_i}^{\phi_f} \frac{V}{V'} d\phi \tag{1.116}$$

with ϕ_i and ϕ_f being the values of the scalar field at the beginning and at the end of the inflationary stage. The last equality of eq. (1.116) is obtained using the equation of motion for the scalar field (1.109) and the expression of the Hubble expansion rate (1.10) at the epoch of inflation. From eq. (1.116) it is clear that the amount of inflation generated is deeply entangled with the shape of the inflaton field potential. We can also estimate the distance that a light ray may travel during the time interval that inflation lasts,

$$d\left(t_i, t_f\right) = a\left(t_f\right) \int_{t_i}^{t_f} \frac{dt}{a\left(t\right)} = H^{-1}\left(e^{H\left(t_f - t_i\right)} - 1\right) \simeq \frac{e^{\mathcal{N}}}{H} \tag{1.117}$$

The distance given by eq. (1.117) shows that a causally-connected region with size equal to the Hubble volume is exponentially expanded by $e^{\mathcal{N}}$ at the end of inflation, whereas the Hubble radius itself remains approximately constant during the same period, $R_H = H^{-1} \simeq$ const.. The Hubble radius will start to grow again after the end of inflation and eventually will enclose regions which are beyond the Hubble volume before inflation started. This can be rephrased in terms of comoving coordinates. During inflation the comoving Hubble length decreases proportionally to $\backsim e^{-\mathcal{N}}$. Comoving scales of the size of the Hubble radius and smaller are therefore pushed outside the comoving Hubble sphere. With the end of inflation, these scales will re-enter progressively the Hubble volume as the comoving Hubble radius starts to increase. If the number of e-foldings is sufficiently large, scales that didn't have time to establish causal contact before inflation are still today beyond our observable horizon. This would explain the high degree of isotropy and homogeneity we observe today. This can also explain why magnetic monopoles and other Big Bang relics, predicted by particle physics, are unobserved up to day. In the case that inflation occurs before or during the phase when such particles are generated, at the end of the inflationary stage, their number density will decrease by a maximum factor of $\backsim e^{3\mathcal{N}}$. If the number of e-foldings $\mathcal{N}$ is large enough, thus the density of these particles will be very small at present time and therefore, we are unable to detect them observationally. The later only work, however, if inflation last a long time, enough to dilute the unwanted relics (see [55] for a discussion on the conditions required for this argument to hold for different kinds of relics). The minimum amount e-folding needed to solve the previous and other problems of the standard big-bang model is about $\mathcal{N} \backsim 60$ (see e.g. [55, 94]).

Primordial fluctuations from Inflation Even though cosmic inflation is very successful in leading the primordial Universe towards homogeneity and flat geometry, a requirement of the Standard Cosmological Model, perhaps the most outstanding virtue of inflation is that it provides a natural

way of producing the primordial inhomogeneities needed to set the seeds for the formation of the amazing structures observed across the Universe. During the period of inflation, the energy density associated with the inflaton field dominates over the energy content of the Universe, thus being responsible for the accelerated expansion of spacetime. Therefore, a fluctuation $\delta\phi$ in the scalar field implies a perturbation in the energy density field, which will be coupled with the metric perturbations through Einstein's field equations (eq. (1.7)). Quantum fluctuations of the inflation field arise on scales much smaller than the comoving Hubble radius. On these small scales, the time evolution of the perturbations is done resorting to flat spacetime quantum field theory. The accelerated expansion of spacetime that occurs in the inflationary period stretches the wavelength of the fluctuations up to scales outside the Hubble radius, where they follow the classical evolution and have their amplitude "frozen-in" at the value of horizon crossing. The perturbations in the inflaton field will then be inherited by the radiation and matter fields to which the inflaton field decays during reheating phase. Soon after the inflationary period ends, the Hubble radius increases faster than the scale factor. Perturbations will eventually re-enter the Hubble horizon at some point in the radiation or matter dominated era (see [95] for further details), where the ones with larger wavelength will, due to gravitationally instability, collapse and give rise to the structures we observe in the Universe today.

Chapter 2

Primordial non-Gaussianities in the Large Scale Structure

In section 1.6 of the previous chapter, we have introduced the inflationary paradigm as a possible solution for a number of problems of the Standard Cosmological Model. In particular, this framework offers a simple, yet powerful, mechanism for the generation of the primordial density fluctuations needed to seed the structure formation that took place in the Universe throughout its evolution. Most standard single field slow-roll inflationary scenarios with canonical kinetic terms predict the production of nearly Gaussian primordial density fluctuations and thus the level of non-Gaussianity associated with these models is not strong enough to be detected [28, 95, 96]. The observed CMB anisotropies and LSS data seem to support such predictions. However, if at least one of the conditions leading to the standard single-field, slow-roll inflationary scenario fails [97], a significant larger non-Gaussianity amplitude can be produced in more complex inflationary models, increasing the possibility of detection with future high precision experiments.

The conditions of slow-roll are directly related to the steepness of the inflaton field potential. In order for inflation to last for the sufficient amount of e-folds, the potential must be flat enough so that the scalar field rolls down towards the minimum very slowly. However, non-Gaussianity can be generated even without necessarily violating this condition. In fact, if certain conditions are fulfilled [98, 99], multi-field inflation models, where the slow-roll conditions holds, can produce a potentially observable large level of non-Gaussianity. Nonetheless, the creation of primordial non-Gaussianities can also be achieved without focusing on the breaking of the slow-roll regime [100, 101].

The diversity of inflationary models capable of giving rise to large non-Gaussianity include the curvaton scenario [102–104], the ekpyrotic inflationary scenario [105, 106], vector field populated inflation [107–109] and multi-field inflation [101, 110–112]. The full description and study of these models is beyond the scope of this thesis and we refer the reader to [95, 113] for a more detailed review on the subject. Another possible mechanism to generate non-Gaussian perturbations is associated with topological defects which are expected to arise during phase transitions in the early Universe. Although they are excluded as the main source of cosmological perturbations, they are still expected to lead to significant deviations from a Gaussian probability distribution,

mainly on small scales [114]. In particular, the hybrid inflation scenario relies on a phase transition which leads to the production of defects at the end of the inflationary era [115]. Cosmic strings have recently been revived as seeds of cosmic structure by work on brane inflation, in the context of superstring theory, where string production at a significant, yet subdominant level, seems to be an unavoidable outcome [116, 117].

It has been reported the detection of Primordial non-Gaussianities in the properties of the CMB temperature anisotropies at a significant level [67, 118]. Although there are contradicting analyses [34, 119, 120], this has served to ignite (or re-ignite) interest in tests of deviations from Gaussianity in the primordial density perturbations. Detection could be achieved through the statistical characterization of the properties of the large-scale structure, including the bispectrum and/or trispectrum of the galaxy distribution (e.g. [31, 37]), the determination of the evolution with time of the abundance of massive collapsed objects such as galaxy clusters (see e.g. [43–47, 121]) or high-redshift proto-galactic objects.

Thus, detecting and constraining primordial non-Gaussianity has become a crucial task for current cosmological studies, since it would allow us to discriminate between the proposed mechanisms for the generation of the density perturbations responsible for the large scale structure we observe in the Universe. In particular, the detection of Primordial non-Gaussianity, would give us an insight on key processes that took place in the early Universe.

2.1 Primordial non-Gaussianity parametrization

Distinguishing between the different types of inflation models capable of generating large non-Gaussianities, requires more information than the one provided by the power spectrum. Thus accessing the non-Gaussian part of the primordial density perturbations is essential. In the case of Gaussian random fields, all information needed to perform their statistical description is contained in the power spectrum and thus it is the only quantity we need to measure. This is no longer true in the presence of non-Gaussian random fields. Therefore, to gain further information on non-Gaussianity, the measurement of higher order correlation functions (higher than the two-point correlation function) is crucial. The three-point correlation function of the primordial curvature perturbation, ζ, is the lowest order statistics sensitive to non-Gaussian features. Performing a Fourier transform of the three-point correlation function, one is able to define the bispectrum, B_ζ, as

$$\langle \zeta(\mathbf{k}_1)\,\zeta(\mathbf{k}_2)\,\zeta(\mathbf{k}_3)\rangle = (2\pi)^3\,\delta^{(3)}(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\,B_\zeta(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\,. \tag{2.1}$$

Similarly to the three-point function, the bispectrum measures the correlation of fluctuations in three points in Fourier space, forming a triangle with three wave vectors. Therefore, the wealth of cosmological information carried by the bispectrum is clearly far greater than the power spectrum, which only correlates two points is space. Depending on the underlying physical mechanism responsible for generating non-Gaussianities, different shapes for the triangle configuration in the bispectrum will arise. This shape is closely related to the amplitude of the ratios k_2/k_1 and k_3/k_1, provided that the overall momentum $\mathbf{K} = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ remains fixed and restricted to zero (as implied by the

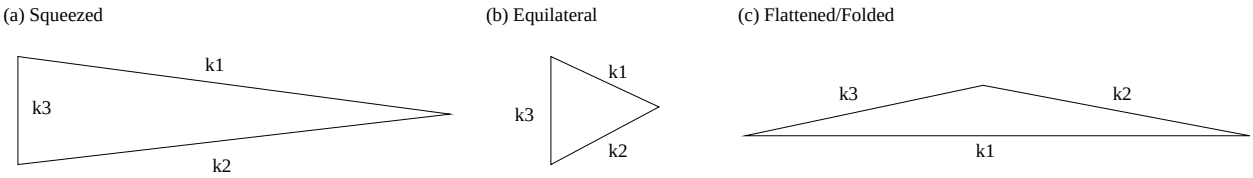


Figure 2.1: The triangle configurations for the different shapes of primordial non-Gaussianities contributing to the bispectrum corresponding to Local (or squeezed) configuration with $k_3 \ll k_1,\ k_2$ (a), Equilateral configuration with $k_3 \approx k_1 \approx k_2$ (b), and Folded/Flattened configuration with $k_3 \approx k_1 + k_2$ (c) [97].

Dirac delta in eq. (2.1)) [46, 122, 123].

There are broadly four classes of triangle shapes that can be found in the literature: Local, Equilateral, Folded and Orthogonal. In figure 2.1 three of the four triangle configurations are shown. The Orthogonal shape is build such that it is orthogonal to both the Local and Equilateral configurations.

It is possible to write the curvature perturbation bispectrum in the form [123],

$$B_\zeta\left(\mathbf{k}_1.\mathbf{k}_2.\mathbf{k}_3\right) = f_{NL}\left(k_1k_2k_3\right)^{-2} A\left(\mathbf{k}_1.\mathbf{k}_2.\mathbf{k}_3\right), \tag{2.2}$$

where A is the shape function, which stores the information about functional dependence of the primordial bispectrum on triangle configuration of the momentum wave vectors. On the other hand, the dimensionless parameter f_{NL} controls the magnitude of each shape [122]. In other words, this parameter quantifies the level of deviation from Gaussian initial conditions and hence the amplitude of non-Gaussinity in each inflationary model, generating the relevant bispectrum shape. Non-Gaussianities of the Local type are the most studied type in the literature. This shape is characteristic of models that develop non-linearities outside the horizon, models where there is an extra contribution to the primordial density fluctuations other than that of inflation, multi-field and curvaton scenarios, inhomogeneous reheating models or ekpyrotic inflation [105]. For this type of non-Gaussianity, the shape function, defined previously in eq. (2.2), has its peak at the squeezed triangle configuration, i.e $k_3 \ll k_1,\ k_2$. On the other hand, non-Gaussianity of the Equilateral type, may be produced by DBI inflation models [124] and also in models with high order derivatives terms. The shape function of the Equilateral type has its peak when the wave vectors form an Equilateral triangle, i.e. $k_3 \approx k_1 \approx k_2$. The Folded shape can be generated if we consider a non-Bunch-Davies vacuum for the initial conditions and in this configuration the maximum of the shape function occurs for $k_3 \approx k_1 + k_2$. Lastly, the peak for the Orthogonal shape occurs when the wave vectors are orthogonal to each other. Therefore, the different inflation models will have quite specific signatures in the shape of the bispectrum.

In the case of the Local type non-Gaussianity it is possible to write the primordial curvature perturbation, $\zeta\left(\mathbf{x}\right)$, as a Taylor expansion around an auxiliary Gaussian random field, ζ_G, as [125, 126]

$$\zeta\left(\mathbf{x}\right) = \zeta_G\left(\mathbf{x}\right) + \frac{3}{5} f_{NL}^{local}\left(\zeta_G^2\left(\mathbf{x}\right) - \left\langle\zeta_G^2\left(\mathbf{x}\right)\right\rangle\right), \tag{2.3}$$

where f_{NL}^{local} controls the level of non-Gaussianity in this configuration (as discussed before). We

can also describe the Local type non-Gaussianity in terms of a perturbation in the primordial gravitational potential, as

$$\Phi\left(\mathbf{x}\right)=\phi\left(\mathbf{x}\right)+f_{NL}^{local}\left(\phi^{2}\left(\mathbf{x}\right)-\left\langle \phi^{2}\left(\mathbf{x}\right)\right\rangle \right), \tag{2.4}$$

where $\Phi\left(\mathbf{x}\right)$ is the gauge-invariant Bardeen potential and $\phi\left(\mathbf{x}\right)$ is a Gaussian random field [126–128].

The Bardeen's potential is related to the primordial curvature perturbation via,

$$\zeta=\frac{5+3w}{3+3w}\Phi+\frac{2}{3\left(1+w\right)H}\dot{\Phi}, \tag{2.5}$$

where we have assumed an equation of state $p=w\rho$ for the description of a perfect fluid. Furthermore, on large scales, in a period of constant w, we have $\dot{\Phi}=0$. Therefore, deep into the era of matter domination, we have

$$\zeta=\frac{5}{3}\Phi. \tag{2.6}$$

Up until now, we have not provided a functional form for the primordial bispectum for each type of primordial non-Gaussianity. For the Local type, which is described mathematically by eq. (2.4), we can write the bispectrum in terms of the primordial power spectrum, $P_{\phi}\left(k\right)$, of the Gaussian primordial gravitational potential as [126],

$$B_{\Phi}^{loc}=2f_{NL}^{local}\left[P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)+P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{3}\right)+P_{\phi}\left(k_{2}\right)P_{\phi}\left(k_{3}\right)\right]. \tag{2.7}$$

In turn, for the Equilateral shape non-Gaussianity, the form of the primordial bispectrum is given instead by,

$$\begin{aligned}B_{\Phi}^{equil}=&6f_{NL}^{equil}\Big[-\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)+2cyc.\right)-2\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)P_{\phi}\left(k_{3}\right)\right)^{2/3}\\&+\left(P_{\phi}\left(k_{1}\right)^{(1/3)}P_{\phi}\left(k_{2}\right)^{(2/3)}P_{\phi}\left(k_{3}\right)+5cyc.\right)\Big].\end{aligned} \tag{2.8}$$

Finally, the bispectrum for the Orthogonal and Folded shape have the form,

$$\begin{aligned}B_{\Phi}^{orthog}=&6f_{NL}^{orthog}\Big[-3\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)+2cyc.\right)-8\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)P_{\phi}\left(k_{3}\right)\right)^{2/3}\\&+3\left(P_{\phi}\left(k_{1}\right)^{(1/3)}P_{\phi}\left(k_{2}\right)^{(2/3)}P_{\phi}\left(k_{3}\right)+5cyc.\right)\Big],\end{aligned} \tag{2.9}$$

and

$$\begin{aligned}B_{\Phi}^{folded}=&6f_{NL}^{folded}\Big[-3\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)+2cyc.\right)+3\left(P_{\phi}\left(k_{1}\right)P_{\phi}\left(k_{2}\right)P_{\phi}\left(k_{3}\right)\right)^{2/3}\\&-\left(P_{\phi}\left(k_{1}\right)^{(1/3)}P_{\phi}\left(k_{2}\right)^{(2/3)}P_{\phi}\left(k_{3}\right)+5cyc.\right)\Big],\end{aligned} \tag{2.10}$$

respectively. In the previous bispectrum definitions, "$cyc.$" refers to cyclical permutations between the three wave vectors $\mathbf{k}_1$, $\mathbf{k}_2$, $\mathbf{k}_3$.

The strongest constraints on primordial non-Gaussianities were obtained by the Planck collab-

oration [35] (see also [34]), $f_{NL}^{local} = 2.5 \pm 5.7$,$f_{NL}^{equil} = -16 \pm 70$, $f_{NL}^{orthog} = -34 \pm 33$, $(1\sigma$ CL$)$ from temperature alone and $f_{NL}^{local} = 0.8 \pm 5$,$f_{NL}^{equil} = -4 \pm 43$, $f_{NL}^{orthog} = -26 \pm 21$ $(1\sigma$ CL$)$ combining temperature and polarization data.

Before concluding this section, it is important to note that primordial non-Gaussianities have been defined in the primordial curvature perturbation (or equivalently in the Bardeen's potential). However, later in this thesis, we will have the need to propagate PNG to the density field instead. This can be achieved by means of the Fourier transform of the Poisson equation,

$$\delta\left(\mathbf{k}, z\right) = \mathcal{M}\left(k, z\right)\zeta\left(\mathbf{k}\right) = \mathcal{M}\left(k, z\right)\left(\frac{3}{5}\right)\Phi\left(\mathbf{k}\right), \tag{2.11}$$

where we have defined the function $\mathcal{M}$ as

$$\mathcal{M} = \frac{2}{5}\frac{1}{H_0^2\Omega_{m0}}T\left(k\right)k^2, \tag{2.12}$$

and $T\left(k\right)$ is the transfer function. There is a diversity of codes that were developed to numerically compute the detailed shape of the transfer function given a particular model of the universe, such as CMBFAST [129–131], CAMB [132], CMBEASY [133] and, more recently, CLASS [134]. The output of these codes is strongly dependent on the physical properties of the mechanism responsible for the perturbations and also on the background model assumed for the universe. For the structure formation models considered in this thesis, we will assume a cold dark matter scenario and, unless otherwise stated, we will make use of the transfer function given by the following fitting formula [135],

$$T\left(q\right) = \frac{\ln\left(1 + 2.34q\right)}{2.34q}\left[1 + 3.89q + \left(16.1q\right)^2 + \left(5.46q\right)^3 + \left(6.71q\right)^4\right]^{-1/4}, \tag{2.13}$$

with $q \equiv k\left(Mpc/h\right)/\Gamma$, and where Γ is the so-called shape parameter [136], given by

$$\Gamma = \Omega_{m0}h\exp\left[-\Omega_{b0}\left(1 + \sqrt{2h}/\Omega_0\right)\right]. \tag{2.14}$$

Eq. (2.13) is the well known BBKS transfer function for CDM modes and was initially derived having in mind a pure cold dark matter scenario ($\Omega_{b0} = 0$). Accounting for the baryonic component does not change the overall shape of eq. (2.13), however it has the effect of lowering the "apparent" dark matter density parameter by the exponential factor in eq. (2.14). Furthermore, in the previous expression the number of families of relativistic neutrinos was assumed to be equal to $N_\nu = 3$. For a different number of neutrino families a multiplicative factor of $\left(g_*/3.36\right)^{-1/2}$ is often introduced in eq. (2.14), where $g_* = 2 + 2\left(7/8\right)\left(4/11\right)^{4/3}N_\nu$, which is the number of effective neutrino degrees of freedom (see e.g. [137]).

Lastly, to compare the theory of structure formation with observations, a smoothing window with a characteristic scale R has to be applied to the density field. Therefore eq. (2.11) smoothed with a filter $W\left(kR\right)$ will acquire the form,

$$\delta\left(\mathbf{k}, z\right) = \mathcal{M}\left(k, z\right)W\left(kR\right)\zeta\left(\mathbf{k}\right). \tag{2.15}$$

With the previous definitions in mind, we are now in a position of relating the bispectrum of the curvature perturbation with the bispectrum of the smoothed density field, which can be written in the form,

$$B_\delta\left(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3\right)=\left(\sum_{i=1}^{3}W\left(k_iR\right)\mathcal{M}\left(k_i,z\right)\right)B_\zeta\left(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3\right). \tag{2.16}$$

The three-point correlation function for the smoothed density field can be easily computed from the bispectrum as,

$$\langle\delta_R^3\rangle=\int\frac{d\mathbf{k}_1}{(2\pi)^3}\int\frac{d\mathbf{k}_2}{(2\pi)^3}\int\frac{d\mathbf{k}_3}{(2\pi)^3}B_\delta\left(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3\right). \tag{2.17}$$

The detection and observational estimation of PNG is of monumental importance, since it could provide an unexplored window to the early Universe. Furthermore, every inflationary model has its unique footprint that is encoded in the bispectrum (or even higher order correlation functions) and thus PNG might serve as a way to discriminate among the different inflationary scenarios. Hence, a positive detection of a non-Gaussian signal in either of the observables used as probes, namely the CMB anisotropy, the abundance of galaxies clusters, the galaxy and galaxy clusters clustering, would take us a step closer towards a better understanding of the physics processes of the very early universe that seeded the structures we are able to observe today.

2.2 The non-Gaussian halo mass function

The Press-Schechter formalism, discussed in section 1.4.6, can be generalized in order to incorporate the effects of PNG in the mass function of the dark matter halos. In fact, the distribution of peaks of the density field where the halos eventually collapse will be strongly dependent on the existence of non-Gaussianities in the primordial curvature perturbations.

To model the deviations from Gaussianity in the mass function we have to obtain an expression for the PDF that accounts for this feature. We do not know a priori the shape of the PDF. However, in general, it is possible to construct the PDF from its cumulants. To see this we follow the approach of [46] (see also [138, 139]) and we start by defining the moments, m_J, of the PDF for the smoothed density field, δ_R, similarly to eq. (1.74),

$$m_J\equiv\left\langle\delta_R^J\right\rangle=\int p\left(\delta_R\right)\delta_R^J d\delta_R, \tag{2.18}$$

where the integer J denotes the order of the corresponding moment. In a similar way, we introduce the cumulants, K_J ,

$$\mathsf{K}_J\equiv\left\langle\delta_R^J\right\rangle_c=\left.\frac{d^J\ln\left[\mathcal{M}\left(y\right)\right]}{dy^J}\right|_{y=0}. \tag{2.19}$$

The function $\mathcal{M}\left(y\right)$ in eq. (2.19) is often referred as the *moment generating function* and is defined

in terms of the moments of the PDF,

$$\mathcal{M}(y) \equiv \langle e^{y\alpha} \rangle = \int e^{y} p(\alpha)\, d\alpha = \sum_{J=0}^{\infty} \frac{y^J}{J!} m_J. \tag{2.20}$$

Finally, the PDF for δ_R is obtained by performing the inverse Laplace transform of eq. (2.20),

$$p(\delta_R) d\delta_R = \frac{d\delta_R}{2\pi i} \int_{-i\infty}^{+i\infty} dy \, \exp\left(-\mathcal{C}(y) - \delta_R y\right), \tag{2.21}$$

where $\mathcal{C}(t)$ is given by,

$$\mathcal{C}(t) \equiv \sum_{J=2}^{\infty} \frac{\mathsf{K}_J}{J!} t^J. \tag{2.22}$$

Eq. (2.21) allows us to recover the PDF for δ_R, knowing the whole series of its cumulants of the distribution. To continue with the derivation, let us define an auxiliary function $\Psi(y, \sigma_R)$ as,

$$\Psi(y, \sigma_R) = \sum_{p=1}^{\infty} S_p(\sigma_R) \frac{(-1)^{p-1}}{p!} t^p = y - \sigma_R^2 \mathcal{C}\left(-y/\sigma_R^2\right), \tag{2.23}$$

where the reduced cumulants are defined by $S_p(\sigma_R) \equiv \langle \delta_R^p \rangle_c / \sigma_R^{2p-2}$ and σ_R is the variance of smoothed density field δ_R. Hence, eq. (2.21) can be re-written in the form,

$$p(\delta_R) d\delta_R = \frac{d\delta_R}{2\pi i \sigma_R^2} \int_{-i\infty}^{+i\infty} dy \, \exp\left(-\frac{\Psi(y, \sigma_R)}{\sigma_R^2} + \frac{\delta_R y}{\sigma_R^2}\right). \tag{2.24}$$

Applying the saddle point approximation to eq. (2.24) and grouping the terms of the same order in σ_R, we then obtain the Edgeworth expansion for the PDF δ_R

$$p(\delta_R)\, d\delta_R = \frac{d\delta_R}{\left(2\pi\sigma_R^2\right)^{1/2}} \exp\left(-\frac{\nu^2}{2}\right) \left[1 + \sigma_R \frac{S_3}{6} H_3(\nu) + \sigma_R^2 \left(\frac{S_4}{24} H_4(\nu) + \frac{S_3^2}{72} H_6(\nu)\right) + \right.$$
$$\left. + \sigma_R^3 \left(\frac{S_5}{120} H_5(\nu) + \frac{S_3 S_4}{144} H_7(\nu) + \frac{S_3^3}{1296} H_9(\nu)\right) + ...\right], \tag{2.25}$$

where we have set $S_P \equiv S_p(\sigma_R)$, $\nu \equiv \delta_R/\sigma_R$ and H_n are the Hermite polynomials given by,

$$H_n(\nu) \equiv (-1)^n \exp\left(\frac{\nu^2}{2}\right) \frac{d^n}{d\nu^n} \exp\left(-\frac{\nu^2}{2}\right) = n! \sum_{i=0}^{[n/2]} \frac{(-1)^n \nu^{n-2k}}{k!\,(n-2k)!\,2^k}, \tag{2.26}$$

where $[n/2]$ means the largest integer $i \leq [n/2]$ [140].

Finally, combining eq. (1.96) with eq. (2.25), and after some algebraic manipulation, a mathematical expression for the mass function of the collapsed dark matter halos, accounting for the existence of primordial non-Gaussianities can thus be written as [46],

$$\frac{dn(z,M)}{dM} = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M}e^{-\frac{\delta_c^2}{2\sigma_M^2}}\left\{\frac{d\ln\sigma_M}{dM}\left[\frac{\delta_c}{\sigma_M}+\frac{S_3\sigma_M}{3!}\left(\frac{\delta_c^4}{\sigma_M^4}-2\frac{\delta_c^2}{\sigma_M^2}-1\right)\right.\right.$$
$$+\frac{1}{2}\left(\frac{S_3\sigma_M}{3!}\right)^2\left(\frac{\delta_c^7}{\sigma_M^7}-13\frac{\delta_c^5}{\sigma_M^5}+25\frac{\delta_c^3}{\sigma_M^3}+15\frac{\delta_c}{\sigma_M}\right)$$
$$\left.+\frac{S_4\sigma_M^2}{4!}\left(\frac{\delta_c^5}{\sigma_M^5}-4\frac{\delta_c^3}{\sigma_M^3}-3\frac{\delta_c}{\sigma_M}\right)\right]$$
$$+\frac{\sigma_M}{6}\frac{dS_3}{dM}\left(\left(\frac{\delta_c^2}{\sigma_M^2}-1\right)+\frac{S_3\sigma_M}{3!}\frac{\delta_c}{\sigma_M}\left(\frac{\delta_c^4}{\sigma_M^4}-10\frac{\delta_c^2}{\sigma_M^2}+15\right)\right)$$
$$\left.+\frac{\sigma_M^2}{4!}\frac{dS_4}{dM}\frac{\delta_c}{\sigma_M}\left(\frac{\delta_c^2}{\sigma_M^2}-3\right)+...\right\}. \tag{2.27}$$

The previous expression for the mass function has a closed form, provided that we know the full series of the cumulants for δ_R. In practice, however, we may only have access to a few low order cumulants of the distribution, such as the variance, skewness and the kurtosis. In this situation, we are forced to use an approximate form for the mass function, and eq. (2.27) will inevitably have a limited validity range (see [46] for a discussion). This range of validity will of course depend on the mass M (since σ_M decreases with M), on the redshift (since the collapse threshold δ_c is larger a high redshift) and of course on the level of non-Gaussianity, i.e., the values of f_{NL}.

For the purpose of the work to be carried out later in this thesis, it is sufficient to only keep the terms up to $\sigma_M S_3$ in eq. (2.27). Hence the mass function will reduce to,

$$\frac{dn}{dM}(M,z) = -\sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}e^{-\delta_c^2(z)/2\sigma_M^2}\left[\frac{d\ln\sigma_M}{d\ln M}\left(\frac{\delta_c(z)}{\sigma_M}+\frac{S_{3M}\sigma_M}{6}\right.\right.$$
$$\left.\left.\times\left(\frac{\delta_c^4(z)}{\sigma_M^4}-2\frac{\delta_c^2(z)}{\sigma_M^2}-1\right)\right)+\frac{1}{6}\frac{dS_{3M}}{d\ln M}\sigma_M\left(\frac{\delta_c^2(z)}{\sigma_M^2}-1\right)\right], \tag{2.28}$$

where S_{3M} is the skewness of the smoothed density field [if $f_{NL}=0$, then $S_{3M}=0$ and eq. (2.28) reduces to the Gaussian halo mass function]. As mentioned in subsection 1.4.6, the mass function derived from the Press-Schechter formalism tends to under-predict the abundance of high mass halos and over-predict the low-mass ones, since the assumption of spherical collapse may not be accurate (see [86] for details). Hence, we will use the same prescription as in [46, 141] to calculate the halo mass function in the presence of non-Gaussianities in the probability distribution of the primordial density perturbations,

$$\frac{dn_{NG}}{dM}(z,M,f_{NL}) = \frac{dn_{ST}}{dM}\frac{dn/dM(z,M,f_{NL})}{dn/dM(z,M,f_{NL}=0)}. \tag{2.29}$$

where dn_{ST}/dM is Sheth and Tormen mass function given in eq. (1.98). Figure 2.2 shows the last term of the right side of eq. (2.29), denominated in the plot as $\mathcal{R}_{NG}$, for different values of the non-linear parameter f_{NL} of the Local type non-Gaussianity, as a function of mass and for two distinct values of redshift. The effects of positive/negative values of f_{NL}, will result in a higher/lower

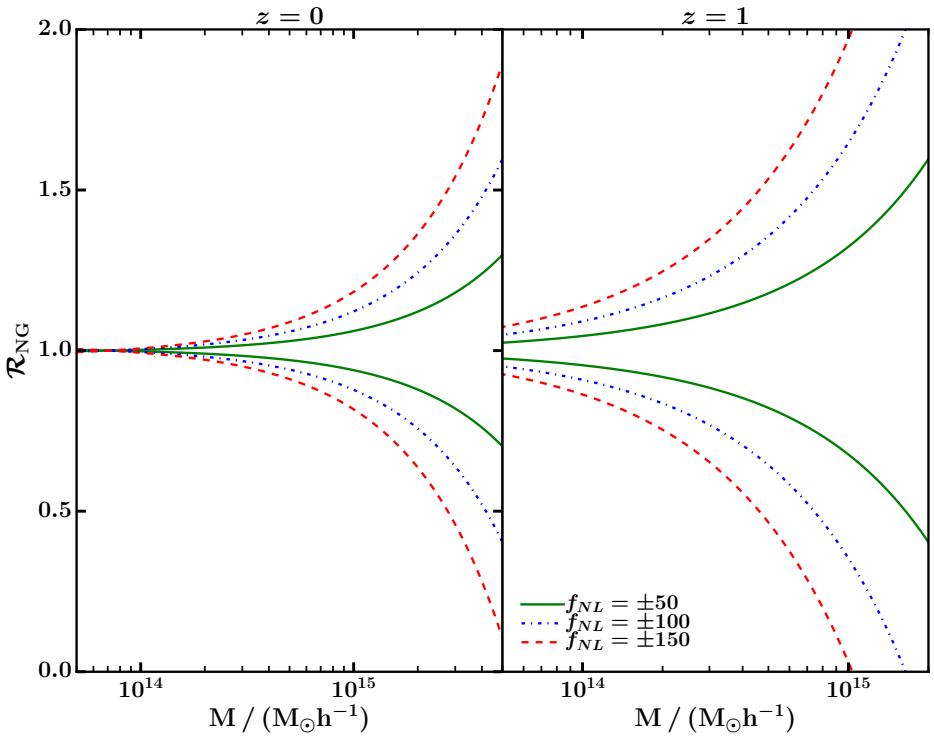


Figure 2.2: The ratio between the non-Gaussian and the Gaussian mass functions as a function of the mass, for different levels of non-Gaussianity. The left panel shows the ratio at redshift $z = 0$, while the right panel shows the ratio computed at the redshift $z = 1$.

number of clusters relative to the Gaussian case, which become more important at high mass and redshift.

Before concluding this section, it is important to point out that eq. (2.27) is not the only prescription to account for possible deviations from Gaussianity in the mass function. However, a thorough discussion on this subject is beyond the scope of this thesis and we refer the reader to [43, 142–147] for more details.

2.3 A new signature of primordial non-Gaussianities from the abundance of galaxy clusters

Galaxy clusters are the largest virialized and gravitationally bounded structures that can be found in the Universe. They constitute a unique and valuable source of cosmological information indispensable in Modern Cosmology, allowing to probe scales much smaller than the ones accessible with current CMB anisotropies measurements. Moreover, the kind of systematic effects that galaxy clusters are subject to are quite different from the ones that affect CMB measurements. The abundance of clusters as a function of mass and redshift contains information that can be used to compare cosmological models or to constrain the value of some parameters associated with them. This observable is sensitive both to the volume of space and to the growth of structure on scales of the order of a few tens of Mpc, as a function of redshift, more specifically, to the expansion his-

tory of the Universe and to the amplitude of the primordial density perturbations on those scales. However, even slight deviations from Gaussianity in the probability distribution of the primordial density perturbations can have a measurable impact on the galaxy cluster abundance, especially at high redshift and masses (see for example [31, 37] and references therein).

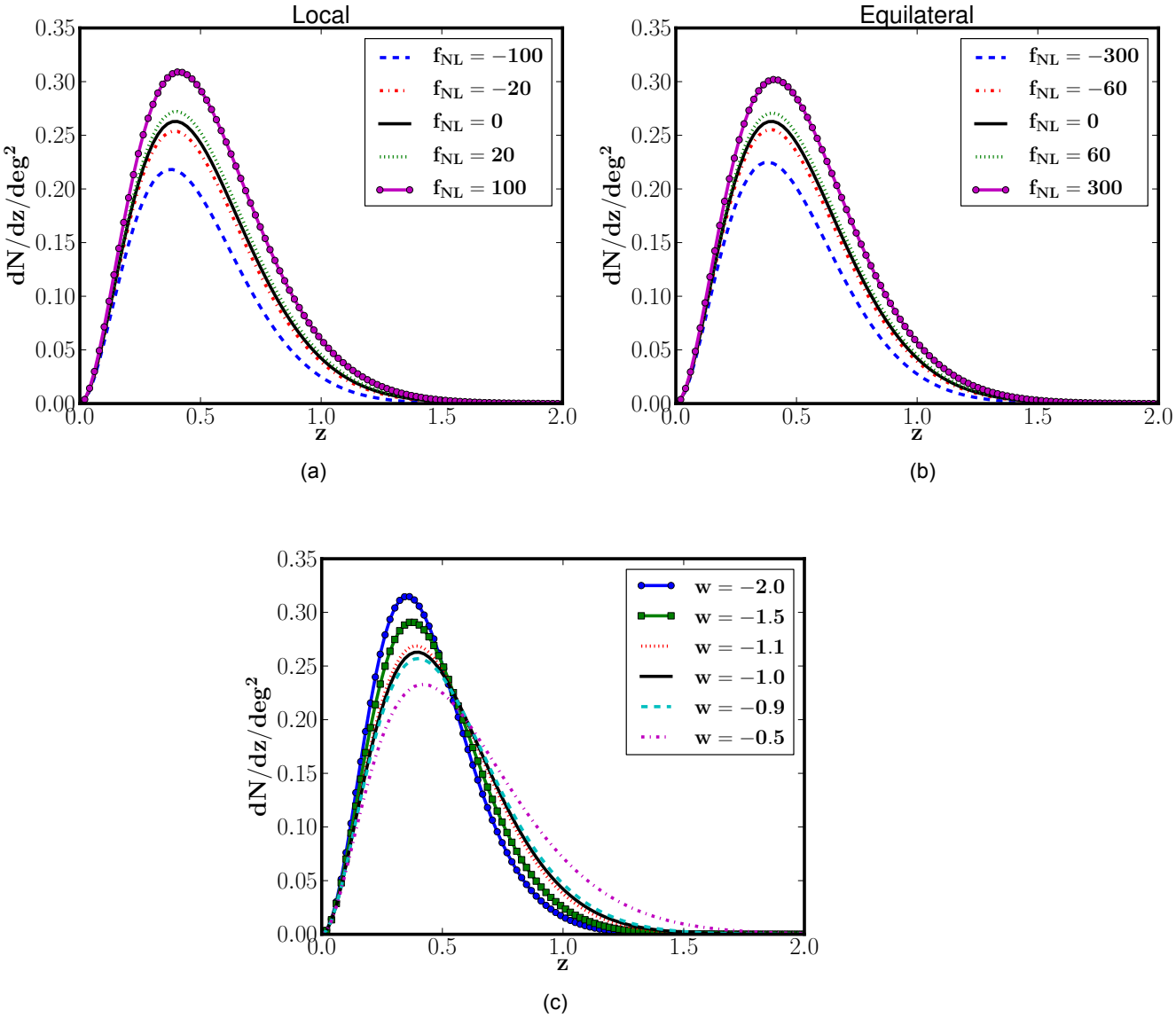


Figure 2.3: The number of galaxy clusters per unit of redshift per square degree with mass $M > M_{lim} = 5{\times}10^{14}\,h^{-1}M_{\odot}$, considering $w = -1$ and different levels of non-Gaussianity for (a) Local and (b) Equilateral parametrizations. Panel (c) shows the effect that a change on a constant dark energy equation of state parameter, w, has on the number of clusters per unit of redshift per square degree with Gaussian initial conditions ($f_{NL} = 0$).

In figures 2.3a and 2.3b, we illustrate the impact of the primordial non-Gaussianities on the number density of galaxy clusters as function of redshift. The curves were computed for the Local and Equilateral parametrizations, respectively (see section 2.1), by combining the non-Gaussian mass function in eq. (2.29) and eq. (1.99), and assuming a mass threshold $M > M_{lim} = 5 \times$

$10^{14}\,h^{-1}M_{\odot}$. Moreover, the cosmological model used was set to be a flat ΛCDM model with WMAP 7-year cosmological parameters (WMAP+BAO+H_0) [32], namely, a Hubble constant, H_0, equal to $100h$ kms^{-1}Mpc^{-1} with $h = 0.704$, fractional densities of matter and baryons today of $\Omega_{m0} = 0.272$, $\Omega_{b0}h^2 = 0.023$ respectively, a scalar spectral index, n_s , equal to 0.963, and we normalize the power spectrum so that $\sigma_8 = 0.809$. Unless otherwise stated, the previous set of cosmological parameters will serve as the fiducial model in the remaining of this section. In figure 2.3c, we have also plotted the number of clusters assuming Gaussian initial conditions (i.e. $f_{NL} = 0$), for the same fiducial cosmological model and mass threshold. However, we have allowed the dark energy equation of state parameter to vary, in order to assess how w, assumed to be constant with time, affects the cluster number density. From figures 2.3a, 2.4b and 2.3c we can see that the effect of changing f_{NL} and w on the abundance of galaxy clusters is quite different. On one hand, increasing w above $w = -1$ flattens the slope of the cluster abundance above $z \approx 0.5$, which translates in an increase in the number of high-z clusters. On the other hand, changes in f_{NL} modify the cluster abundance more uniformly in redshift. The difference in behaviour occurs because w affects both the volume factor and the mass function, while f_{NL} changes only the tail of the distribution of the density perturbations, thus modifying just the mass function.

The reconstruction of w_{eff}

Figures 2.3a, 2.3b and 2.3c, suggest that the redshift evolution of the number density of galaxy clusters in non-Gaussian models could be wrongly taken to be the result of an effective dark energy equation of state different from the real one, under the assumption of Gaussian initial conditions. In order to test this hypothesis, we have generated mock catalogues with the expected redshift evolution of the cluster number density for different non-Gaussian initial conditions in bins of redshift centered at redshift z_i and with width $\Delta z = 0.1$ up to redshift 2. The expected number of clusters in each redshift bin was computed as,

$$N\left(z_i \pm \Delta z\right) = f_{sky} \int_{z_i - \Delta z/2}^{z_i + \Delta z/2} \frac{dV}{dz} \left[\int_{M_{lim}(z)}^{\infty} dM \frac{dn}{dM} \right], \tag{2.30}$$

with the mass function given by eq. (2.29). Furthermore, we have assumed a mass threshold of $M_{lim} = 5 \times 10^{14} h^{-1} M_{\odot}$, a nearly full sky survey area of 40,000 square degrees and we consider $f_{NL}^{local} = (-20, 20)$ and $f_{NL}^{equil} = (-60, 60)$, respectively for the Local and Equilateral parametrizations mentioned in section 2.1. The reason for choosing these specific values for f_{NL} will become clear later.

Having generated the mock catalogues with the expected redshift distribution of the number density of galaxy clusters with non-Gaussian initial conditions, we then computed an effective dark energy equation of state, w_{eff}, using eq. (2.30) with $f_{NL} = 0$, that mimics the distribution of the number of galaxy clusters in the presence of non-Gaussian initial conditions at the *i-th* redshift bin, by solving the following equation,

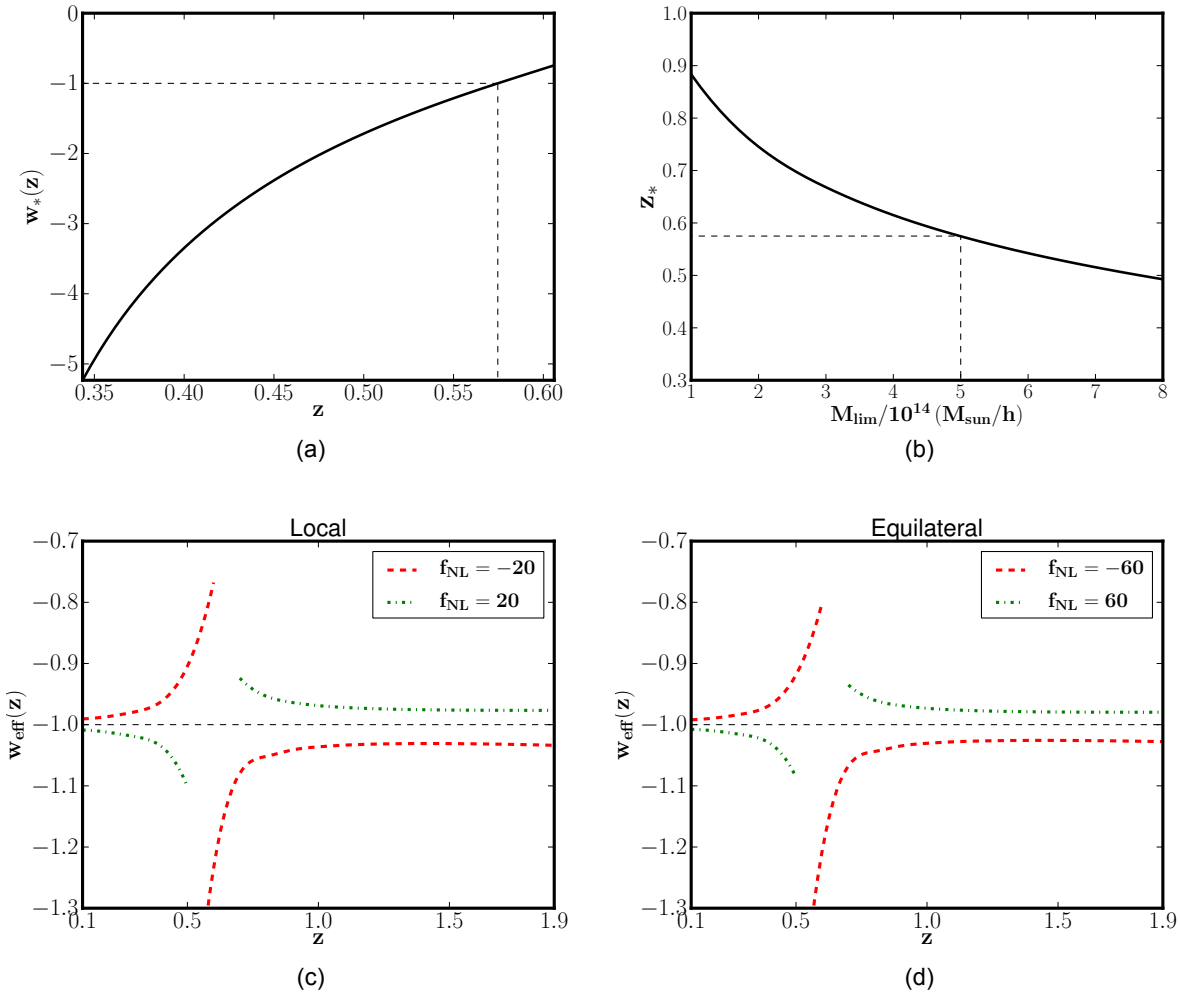


Figure 2.4: Figure 2.4a shows the value of w_{eff} that maximizes the abundance of clusters as a function of redshift, with Gaussian initial conditions and $M_{lim} = 5 \times 10^{14}\, h^{-1} M_\odot$ (the vertical dashed line corresponds to the redshift $z_* \sim 0.575$ where $w_* = -1$). Figure 2.4b shows the dependence of the value of z_* on the mass threshold M_{lim}. Figures 2.4c and 2.4d show the reconstructed effective dark energy equation of state w_{eff} for $M > M_{\text{lim}} = 5 \times 10^{14}\, h^{-1} M_\odot$ and different values of f_{NL} (Local and Equilateral parametrizations, respectively).

$$N_i^G = \int_{z_i-\Delta z/2}^{z_i+\Delta z/2} \frac{dN}{dz}(z, w = w_{eff}, f_{NL} = 0)\, dz = \\ = \int_{z_i-\Delta z/2}^{z_i+\Delta z/2} \frac{dN}{dz}(z, w = -1, f_{NL} \neq 0)\, dz = N_i^{NG}, \tag{2.31}$$

with respect to w_{eff} for each redshift bin. Thus, the effective dark energy equation of state, $w_{eff}(z)$, is defined as the value of w which reproduces the number of clusters of the mock non-Gaussian catalogue in the *i-th* redshift bin.

Model	Local		Equilateral	
f_{NL}	-20	20	-60	60
$\sigma_{8,NG}$	0.812	0.806	0.813	0.807

Table 2.1: The computed σ_8 for the Local parametrization and Equilateral parametrization, obtained by demanding that the present-day number density of galaxy clusters is recovered.

The normalization of the power-spectrum, for models with non-Gaussian initial conditions, was done by demanding that the present-day number density of galaxy clusters in the concordance model (ΛCDM cosmology with Gaussian initial conditions and $\sigma_8 = 0.809$) is recovered. Table 2.1 shows the computed σ_8 for different values of f_{NL} for the Local and Equilateral parametrizations.

Figures 2.4c and 2.4d show the reconstructed dark energy equation of state, w_{eff}, as a function of the redshift for $M_{lim} = 5 \times 10^{14}\, h^{-1} M_\odot$ and different values of f_{NL} (Local and Equilateral parametrizations, respectively). At low redshift, our reconstructed w_{eff} is very close to our fiducial $w = -1$. But, as we move towards higher redshifts, our computed w_{eff} deviates from -1, with this effect being more evident for higher values of f_{NL} in both parametrizations. Further, for $f_{NL} > 0$ there is a redshift interval where w_{eff}, is undefined, which widens with increasing f_{NL}. For $M_{lim} = 5 \times 10^{14}\, h^{-1} M_\odot$ this happens in a redshift range centered at the redshift $z_* \sim 0.575$, at which the value of w_{eff}, which we will call w_*, that maximizes the cluster abundance is equal to -1 (see figure 2.4a). In this interval, there is no w_{eff} that solves eq. (2.31), since the product of the comoving volume with the integral of the mass function, when assuming Gaussian initial conditions, is always smaller than the same quantity for non-Gaussian initial conditions, i.e. $f_{NL} \neq 0$. If our fiducial cosmological model had a different value for w, then the discontinuity would appear at the redshift at which $w = w_*$. On the other hand, for models with $f_{NL} < 0$, we have also a discontinuous w_{eff} but there are two values of w_{eff} capable of reproducing the number of clusters in non-Gaussian models in a small redshift interval centered at $z_\star$. The dependence of the value of z_* on the mass threshold M_{lim} is shown in figure 2.4b.

Estimation of w_{eff} with statistical uncertainties

The observational estimation of the number density of galaxy clusters is affected by two sources of statistical uncertainty: the shot-noise and the cosmic variance (see [148]). The statistical uncertainty associated with the former increases, for example, with the cluster mass threshold, as clus-

ters then become more rare, while the statistical uncertainty associated with the latter increases, for example, as the cosmic volume surveyed gets smaller. However, assuming primordial Gaussian density perturbations, and for a cluster mass threshold above $5 \times 10^{14} h^{-1} M_{\odot}$, it has been shown that the magnitude of the contribution of the cosmic variance to the statistical uncertainty, in the observed number density of galaxy clusters, is always at least an order of magnitude smaller than the contribution due to shot-noise, almost independently of the surveyed sky area (e.g. see figures 5 and 9 of [148]).

The existence of statistical uncertainties in the observed number density of galaxy clusters will to some extent mask the apparent discontinuity on the evolution of w_{eff} described previously , which could be used to identify the presence of non-Gaussian density perturbations. In order to determine the impact of such uncertainties, we will re-estimate w_{eff} in similar fashion to what was done before , but with the inclusion of shot-noise, $\sigma_N = \sqrt{N_i}$, at the 1σ level in each redshift bin. We are able to neglect the contribution from cosmic variance by setting the cluster mass threshold to $5 \times 10^{14} h^{-1} M_{\odot}$, and noting that the assumed level of non-Gaussianity is relatively small, not affecting much the average number density of galaxy clusters with respect to the Gaussian case (see figure 2.3).

Figures 2.5 and 2.6 show the 1σ statistical uncertainty in the reconstructed w_{eff}, as a function of redshift, when shot-noise is taken into account, and as before for a nearly full sky survey area of 40,000 square degrees. As can be seen, even when including the effect of such uncertainty, the apparent discontinuity on the evolution of w_{eff} found in the previous subsection is still present (at least at the 2σ confidence level) for values of f_{NL} as small as ± 20 for the Local parametrization, and ± 60 for the Equilateral parametrization. Clearly, decreasing the survey area would mean that only values of f_{NL} with larger modulus could be detected.

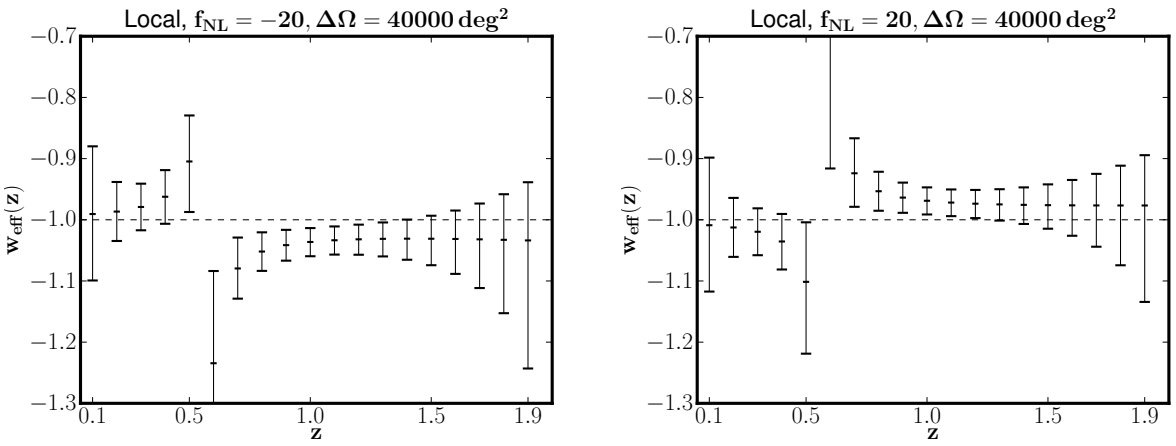


Figure 2.5: The reconstructed effective dark energy equation of state for the Local parametrization with statistical observational uncertainties taken into account.

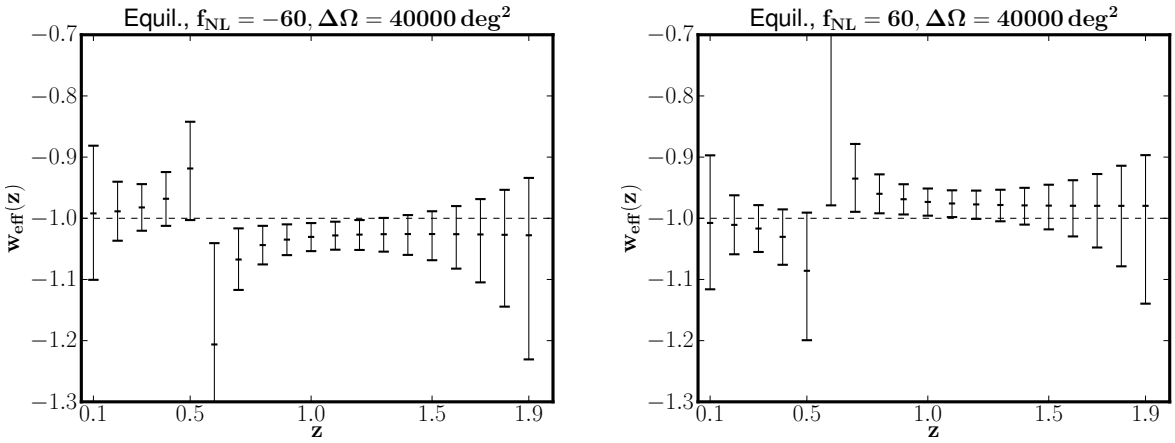


Figure 2.6: The same as in figure. 2.5, but for the Equilateral parametrization.

2.4 Biased cosmological parameter estimation with galaxy cluster counts in the presence of primordial non-Gaussianities

In section 2.3 (see also [141]), we have shown that assuming the absence of primordial non-Gaussianities may lead to an apparent discontinuity in the evolution of the estimated effective equation of state parameter with redshift using galaxy cluster counts. In the same spirit, we will estimate and quantify the biases which might be introduced in the determination of several of the most important cosmological parameters using the evolution with time of the galaxy cluster abundance, if it was wrongly assumed that the initial conditions of the primordial density field were Gaussian distributed. We focus our attention on the present-day dark energy density, Ω_{de0}, the dark energy equation of state parameter w (here assumed to be a constant), and the present-day root mean square mass perturbations, σ_8, at the standard $8\,h^{-1}$ Mpc scale. In this section we will focus our attention mostly on non-Gaussianities of the Local type, although we will also mention at some point the Equilateral configuration.

In order to quantify the biases which may arise in the estimation of cosmological parameters using the redshift evolution of the galaxy cluster abundance due to wrongly assuming the absence of primordial non-Gaussianities, we will follow the same approach considered in [149]. Therefore, we will assume that the likelihood, L, of observing a given number of clusters, for a certain combination of cosmological parameters, in each bin of redshift and mass is given by

$$\ln \mathrm{L} = \sum_{i=1}^{n_z} \sum_{j=1}^{n_M} \left[N_{ij}^m \ln N_{ij}^f - N_{ij}^f - \ln \Gamma \left(N_{ij}^f + 1 \right) \right] , \tag{2.32}$$

where Γ is the Gamma function, n_z and n_M are respectively the number of redshift and mass bins, while N_{ij}^f and N_{ij}^m are the number of counts for respectively the fiducial combination of cosmological parameters and the combination assumed to be true, for the i-th redshift bin and the j-th mass bin. Both N_{ij}^f and N_{ij}^m are computed using eq. (2.30). Our assumed observed cluster catalogue will not be a particular realization of the cluster redshift distribution in the context of the fiducial model, but

a sort of average-catalogue. Nevertheless, this will allow for a very good estimate of the size and shape of the expected likelihood contours expected from an analysis of the information contained in a real cluster catalogue, and avoid the offset in the best fit away from the fiducial values for the cosmological parameters, that would result from using a randomly-generated cluster catalogue based on the fiducial model (for more details see [149]).

In the above expression for the likelihood, we assumed that all clusters are randomly distributed in space, i.e. their positions follow a Poisson distribution. However, those positions are in fact spatially correlated. The effect of these correlations on the cluster abundance is often referred to as cosmic variance, while the variations in the cluster abundance due to the cluster distribution being a Poisson process are known as shot-noise. As already mentioned, assuming primordial Gaussian density perturbations, the contribution of the cosmic variance to the statistical uncertainty associated with the galaxy cluster abundance is at least an order of magnitude smaller than the contribution due to shot-noise, almost independently of the surveyed sky area, as long as the cluster mass threshold is above $5 \times 10^{14} h^{-1} M_{\odot}$ [148]. Therefore, so that we can safely neglect the contribution from cosmic variance, we will set again the cluster mass threshold to $5 \times 10^{14} h^{-1} M_{\odot}$, noting that the level of non-Gaussianity we are assuming is relatively small, not affecting much the galaxy cluster abundance with respect to the Gaussian case [141].

We assume a flat ΛCDM fiducial cosmology (i.e. the fiducial value for w is assumed to be -1), as derived in [14], namely, a Hubble constant, H_0, equal to $100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ with $h = 0.673$, fractional energy densities associated with dark energy and baryons today of $\Omega_{de0} = 0.685$ and $\Omega_{b0}h^2 = 0.02205$ respectively, a scalar spectral index, n_s, equal to 0.9603, and power spectrum normalization of $\sigma_8 = 0.829$. The level of non-Gaussianity is parametrized by the f_{NL} parameter, which we allow to vary from -100 up to 100 with increments of 50. We also consider a fiducial sky area of $4000\,\mathrm{deg}^2$ and we generate the expected number of clusters in redshift bins with width $\Delta z = 0.1$ up to a redshift of 1.4. The cosmological parameters whose values we attempted to constrain using the information contained on those catalogues were Ω_{de0}, w and σ_8, where w is the (assumed independent of redshift) constant equation of state associated with the dark energy. Flat priors were associated with all, rendering their posterior probabilities proportional to the likelihood given by eq. (2.32). We set $f_{NL} = 0$ as a prior, given our objective of quantifying the biases which may arise due to wrongly assuming the absence of primordial non-Gaussianities. The exploration of the likelihood in the defined parameter space was carried out using a custom code based on standard Monte Carlo Markov Chain techniques (e.g. [150]).

As can be seen in figure 2.7, the biases on Ω_{de0} and w, that arise from wrongly assuming f_{NL} to be zero, are very small. In fact, it would be necessary larger values of $|f_{NL}|$ than the ones considered here, combined with multiple mass bins and a larger sky coverage for the fiducial values of Ω_{de0} and/or w to fall outside the derived 2σ confidence levels.

Contrary to Ω_{de0} and w, the bias on the estimation of σ_8 due to wrongly assuming Gaussianity is significantly more severe, although even in this case the exclusion of the fiducial value at more than 2σ requires a value for f_{NL} lower than about -80, or significantly higher than 100. The same exclusion level is attained for lower $|f_{NL}|$ if multiple mass bins are considered and/or the sky area coverage increased. In the later case, we have confirmed that, as expected, the uncertainty as-

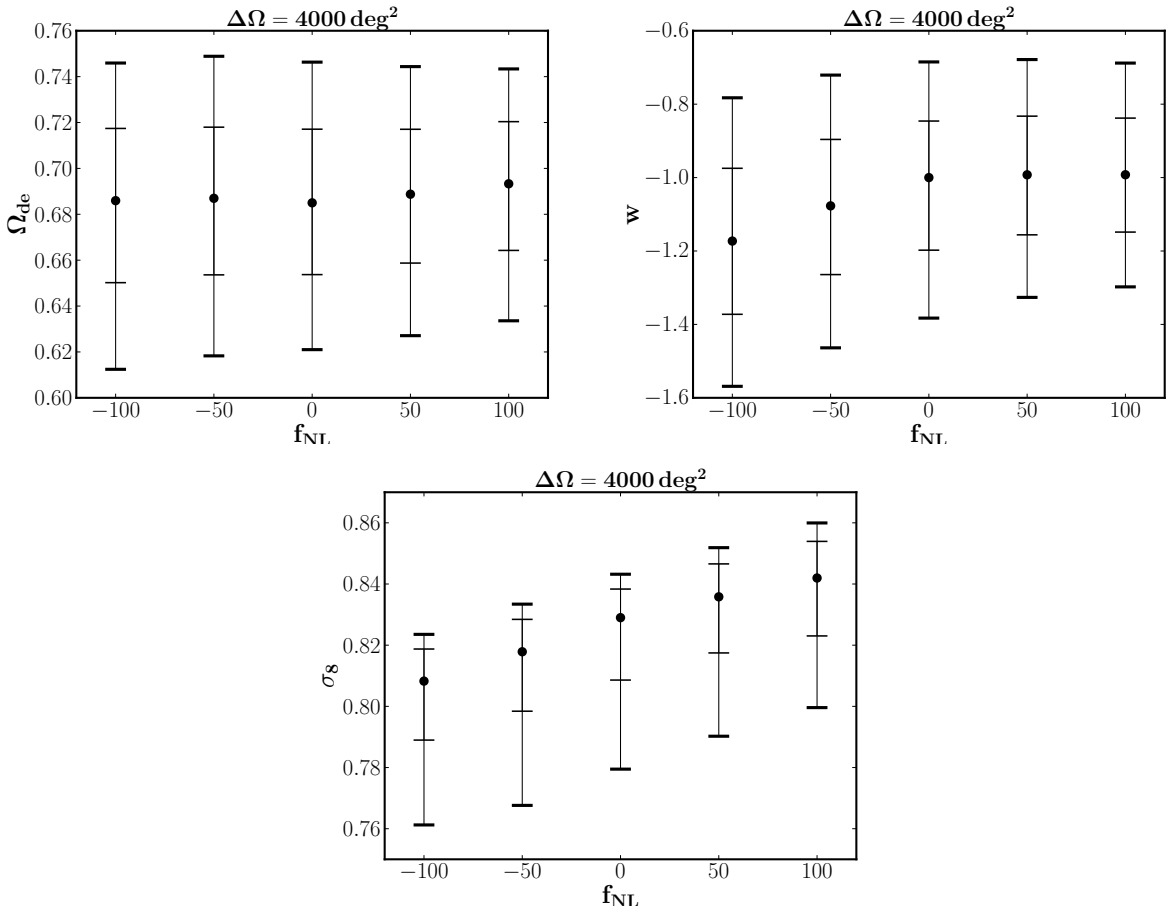


Figure 2.7: The most probable values of Ω_{de0}, w and σ_8, and associated 1σ (inner thin ticks) and 2σ (outer thick ticks) confidence levels, for a sky coverage of $4000\,\mathrm{deg}^2$.

sociated with the estimation of each parameter is inversely proportional to the square root of the sky area coverage. For example, increasing it to $40000\,\mathrm{deg}^2$ would result in the fiducial value for σ_8 assumed here to be excluded at more than 2σ for f_{NL} smaller than -30.

The best fit values of Ω_{de}, w and σ_8 as a function of the non-Gaussianity level are given approximately by

$$\Omega_{de0} = 0.686 + 3.3 \times 10^{-5} f_{NL} + 3.7 \times 10^{-7} \left(f_{NL}\right)^2 , \tag{2.33}$$

$$w = -1.010 + 8.9 \times 10^{-4} f_{NL} - 7.4 \times 10^{-6} \left(f_{NL}\right)^2 , \tag{2.34}$$

$$\sigma_8 = 0.829 + 1.7 \times 10^{-4} f_{NL} - 3.2 \times 10^{-7} \left(f_{NL}\right)^2 , \tag{2.35}$$

with a maximum fitting error below 1%. The induced systematic errors in the cosmological parameters due to a systematic error in the non-Gaussianity parameter can be obtained by simply differentiating eqs. (2.33 - 2.35) with respect to f_{NL}. The dependences expressed through eqs. (2.33 - 2.35) do not change significantly if the cluster mass threshold, assumed here to be $5 \times 10^{14} h^{-1} M_{\odot}$, is changed.

Although the results above are specifically for the Local parametrization of f_{NL}, they can be

used to infer the what would happen had we considered the Equilateral parametrization. In fact, we have found that the abundance of galaxy clusters can also be computed for the Equilateral parametrization using

$$f_{NL}^{local} \simeq \frac{f_{NL}^{equil}}{3.6} \,. \tag{2.36}$$

which seems to be consistent with the recent results of [151] and do not change much with cluster mass threshold.

2.5 Effect of Primordial non-Gaussianities on Galaxy Clusters Scaling Relations

On galaxy cluster scales, primordial non-Gaussianities are expected to influence the mass and redshift distribution of cluster abundances. As discussed in previous sections (see also [141, 152]), failing to take in to account the effect of PNG on clusters abundances, can led to biases in the estimation of cosmological parameters when clusters counts are used as cosmological probes. Accurate predictions on how PNG affects clusters abundances require detailed knowledge about the underlying mass function of cluster halos as well as understanding the way the total cluster mass relates to baryon observables.

The impact of primordial non-Gaussianities on the cluster halo mass function has been extensively investigated using a combination of analytical (see e.g. [43, 44, 46, 52, 146, 147, 153]) and numerical N-body (dark matter only) simulation (see e.g. [154–161]) methods. On the other hand, the study of the effect of PNG on cluster baryon observables is hard to model analytically. Hydrodynamic N-body simulations (that model both dark matter and baryons) are the most appropriate tool to follow the evolution of the complex baryon physics acting on inter-galactic (IGM) and intra-cluster medium (ICM) scales during the non-linear evolution of cosmological structure. A first study of the impact of primordial non-Gaussianities on structure formation using hydrodynamic N-body techniques was carried out by [162], that modelled gas chemistry and a number of other gas physical processes in their simulations to study early gas properties, star formation, metal enrichment and the evolution of stellar populations. Later, [163] also applied hydrodynamic simulations with chemistry and radiative gas physics to study the formation and evolution of galaxies within the PNG framework. More recently [164] carried out PNG hydrodynamic simulations, including cooling, star formation, stellar evolution and metal pollution from stellar populations, to study the Sunyaev-Zeld'ovich (SZ) signal, due to the inverse Compton scattering of CMB photons by ionized gas, in galaxy clusters and filamentary structures

Galaxy cluster number counts, e.g. from X-rays or SZ cluster surveys, are known to be a most promising method to constrain deviations from primordial Gaussianity at cluster scales (see eg [165–169] for several cluster survey forecasts). These estimates critically rely on assumptions about the state of the ICM gas atmospheres and on the way their observed properties link with the total cluster mass. The link is usually expressed via galaxy cluster scaling relations that allow to convert mass function estimates into observed number counts. These studies often assume

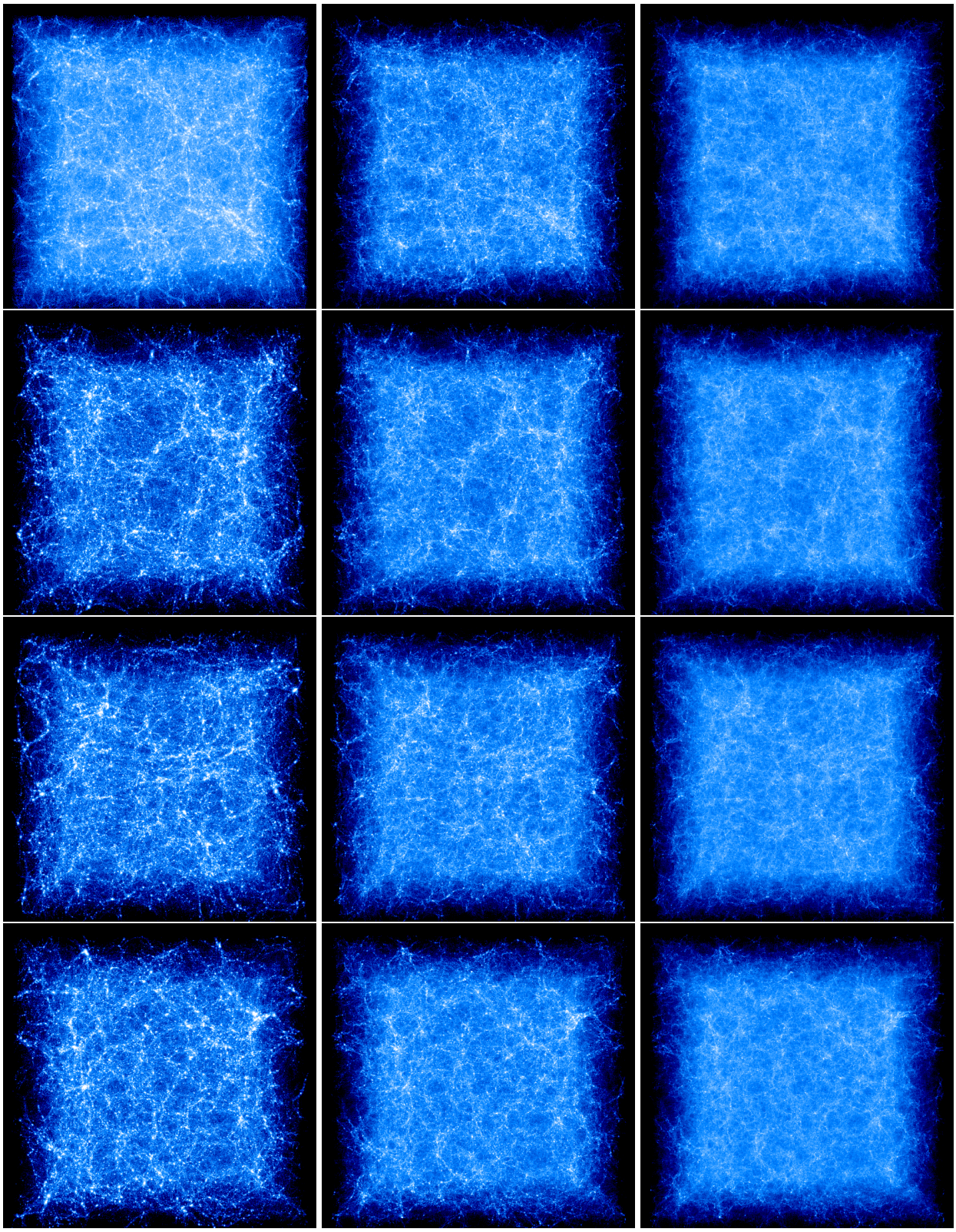
hydrostatic equilibrium, spherical symmetry and the self-similar model for clusters [170, 171]. More sophisticated approaches rely on galaxy cluster scaling relations derived from hydrodynamic or N-body simulations, calibrated by observations, that do not include primordial non-Gaussianities (see eg [166, 168]). This procedure is clearly not ideal given that non-Gaussianities are known to influence the internal structure of clusters [160, 172] and therefore they may cause significant changes in the slope and normalization of galaxy cluster scalings. The study of the impact of PNG on cluster scaling relations is also essential for an accurate characterization of the physical state of the ICM gass and to assess the relative strength of cosmological effects shaping the evolution of galaxy clusters.

In this section, we therefore investigate the effect that primordial non-Gaussianities have on galaxy clusters scaling relations, using hydrodynamic N-body simulations of large scale structure. We focus on scalings involving cluster mass, M, and gas properties related to the thermodynamical state of the intra-cluster medium. These are the temperature, T, entropy, S and the cluster integrated pressure (thermal energy density) expressed by the SZ Y-Compton parameter. Throughout this section, and unless stated otherwise, we adopt a standard flat ΛCDM cosmological model, with a Hubble constant, H_0, equal to $100h$ km/s Mpc^{-1}, with $h = 0.7$, fractional densities of matter and baryons today of $\Omega_{m0} = 0.3$, $\Omega_{b0} = 0.04$ respectively, a scalar spectral index, n_s , equal to 0.96, and a power spectrum amplitude $A = 2.1 \times 10^{-9}$, so that $\sigma_8 = 0.809$.

2.5.1 Numerical Simulations and Catalogue Construction

To assess the impact of primordial non-Gaussianities on galaxy clusters scaling relations, we carried out hydrodynamic N-body simulations of large-scale structure with the publicly available Gadget-2 TreePM code [91], featuring adiabatic gas physics. The simulations initial conditions were generated with the 2LPT code [173], assuming periodic boundary conditions on a cubic volume with $L = 250\,h^{-1}$Mpc on the side and populated with $N = 2 \times 300^3$ particles of baryonic and dark matter. The matter power spectrum transfer function was computed with the CAMB code [132, 174] for the set of cosmological parameters adopted in the previous section. The resulting baryon and dark matter particle masses in the simulations are $6.4{\times}10^{10}\,h^{-1}\mathrm{M}_\odot$ and $4.2{\times}10^{9}\,h^{-1}\mathrm{M}_\odot$, respectively. The gravitational softening in physical coordinates was $30\,h^{-1}$kpc. The initial conditions were generated for different levels of non-Gaussianity of the Local type (see section 2.1), allowing f_{NL} to vary in the range $[-500, 500]$ as indicated in table 2.2. For each value of f_{NL}, 5 random box realizations were created with different seeds, thus resulting in a total of 35 simulation runs. For each run, we have stored a total of 22 snapshots, with abutting boxes, in the redshift range $0 \leq z \leq 2$. In figure 2.7 is shown the output boxes for three values of redshift $z = 0,\, 1,\, 2$ (starting from the most left box to the right) and for increasing values of f_{NL} of table 2.2 (from the top to the bottom of the figure). Doing a visual analysis to the boxes, we can see that the structures tend to be more abundant with ever increasing f_{NL} and z, as one might expect if primordial non-Gaussianities exist.

To construct cluster catalogues for all runs, we used a modified version of the cluster finder software developed by Thomas and collaborators [175–177]. The mass of the identified objects is



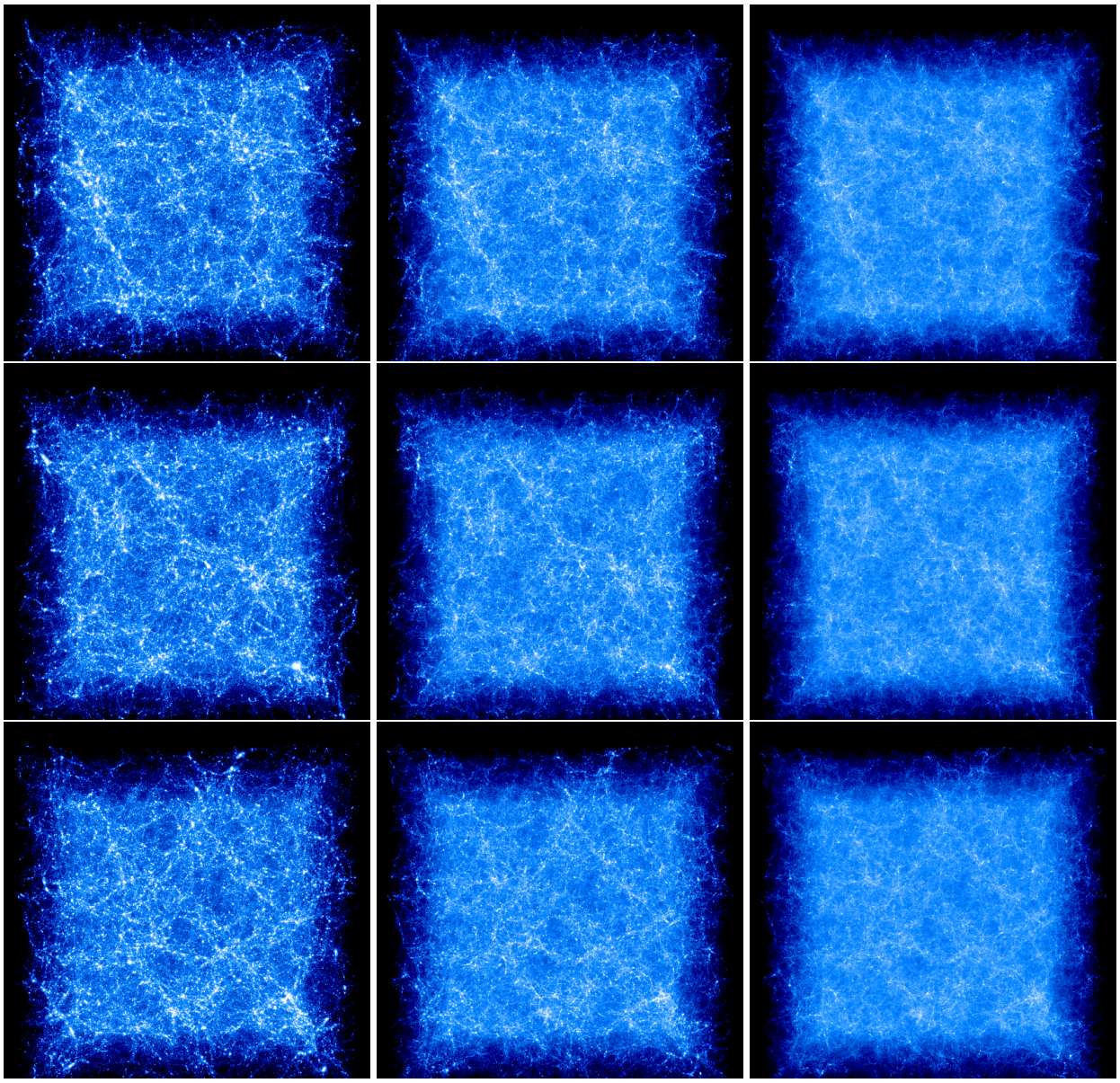


Figure 2.7: The output boxes from the numerical simulations with non-Gaussian initial conditions. Row-wise from the left to the right, it is shown the evolution with the redshift, $z = 0, 1, 2$; while column-wise from the top to the bottom, it is shown the evolution of the boxes as a function of the non-linear parameter f_{NL} with increasing values $-500, -300, -100, 0, 100, 300, 500$.

Table 2.2: List of models considered in this work. Non-Gaussian models are identified by the prefix "NG" followed by the corresponding f_{NL} value, whereas "G" stands for the Gaussian Λ-CDM model. Five initial condition realizations were produced for each model, yielding a total of 35 simulation runs. The quantities $N(z=0)$ and $N(z=1)$ are the total number of clusters when the five realizations for each model are combined. $\overline{N}(z=0)$ and $\overline{N}(z=1)$ give the average number of clusters for each realization of a given model.

Model	f_{NL}	$N\,(z=0)$	$N\,(z=1)$	$\overline{N}\,(z=0)$	$\overline{N}\,(z=1)$
NG-500	-500	3460	771	692	142
NG-300	-300	3499	828	700	166
NG-100	-100	3649	892	730	178
G	0	3653	913	731	183
NG 100	100	3638	925	728	193
NG 300	300	3713	1037	743	207
NG 500	500	3779	1073	756	215

set according to usual definition,

$$M_{\Delta}(< R_{\Delta}) = \frac{4\pi}{3} R_{\Delta}^{3}\, \Delta\, \rho_{\rm crit}(z). \tag{2.37}$$

where Δ is a fixed overdensity contrast, $\rho_{\rm crit}(z) = (3H_0^2/8\pi G)E^2(z)$ is the critical density and $E(z) = H(z)/H_0 = \sqrt{(\Omega(1+z)^3 + \Omega_{\Lambda}}$. Catalogue cluster properties are evaluated inside spheres of radius R_{Δ}, centered around the densest dark matter particle in each cluster. Here we chose $\Delta = 200$ and set the minimum number of cluster particles equal to 500. In this way our original cluster catalogues are complete in mass down to $\approx 3.41 \times 10^{13} h^{-1} M_{\odot}$, at all redshifts. For the present analysis, we trimmed our original catalogues to exclude galaxy groups with masses below $M_{lim} = 5 \times 10^{13} h^{-1} M_{\odot}$. For each model, we also combined catalogues from different realization runs at each redshift to construct single cluster catalogues, all having a minimum mass limit, M_{lim}.

Table 2.2 provides an overview of the number of clusters with masses above M_{lim} at $z = 1$ and $z = 0$ for each of our simulated models. The $N(z=0)$ and $N(z=1)$ are the total number of clusters when the five realizations of each model are combined. $\overline{N}(z=0)$ and $\overline{N}(z=1)$ give the average number of clusters for each realization of a given model. These numbers confirm expectations that cluster abundances are a function of f_{NL}, with negative/positive f_{NL} models giving lower/higher cluster abundances than the Gaussian model, see e.g. [155]. Although our simulations were not set for mass function studies (they have a limited boxsize and five realizations for each model) we see that all our models follow this trend with the exception of the NG 100 model at $z = 0$ that has the largest statistical fluctuation of initial conditions power spectrum amplitudes of all models.

Cluster properties investigated in this thesis are the mass, M, mass-weighted temperature, $T_{\rm mw}$, entropy, S (defined as $S = k_{\rm B} T/n^{-2/3}$ where T and n are the gas temperature and number density), integrated Compton signal, Y (defined as the SZ signal times the square of the angular diameter distance to the cluster), and Y_X, the integrated Compton signal estimated using X-ray emission-weighted temperature, T_X, and the gas mass, M_{gas}. These quantities were computed using their usual definitions, see e.g. [178]:

$$M = \sum_k m_k, \tag{2.38}$$

$$T_{\rm mw} = \frac{\sum_i m_i\, T_i}{\sum_i m_i}, \tag{2.39}$$

$$S = \frac{\sum_i m_i\, k_{\rm B} T_i\, n_i^{2/3}}{\sum_i m_i}, \tag{2.40}$$

$$Y = \frac{k_{\rm B}\sigma_{\rm T}}{m_{\rm e}c^2}\frac{(1+X)}{2m_{\rm H}}\sum_i m_i\, T_i, \tag{2.41}$$

$$Y_X = \frac{k_{\rm B}\sigma_{\rm T}}{m_{\rm e}c^2}\frac{(1+X)}{2m_{\rm H}}\, M_{gas}\, T_{\rm X} \tag{2.42}$$

$$M_{gas} = \sum_i m_i, \tag{2.43}$$

$$T_X = \frac{\sum_i m_i\, \rho_i\, \Lambda_{\rm bol}(T_i, Z)\, T_i}{\sum_i m_i\, \rho_i\, \Lambda_{\rm bol}(T_i, Z)} \tag{2.44}$$

where summations with the index *i* are over hot ($T_i > 10^5$K) gas particles and the summation with the index *k* is over all (baryon and dark matter) particles within R_{200}. Hot gas is assumed fully ionized. The quantities m_i, T_i, n_i and ρ_i are the mass, temperature, number density and mass density of gas particles, respectively. Λ_{bol} is the bolometric cooling function in [179] and Z is the gas metallicity. Other quantities are the Boltzmann constant, $k_{\rm B}$, the Thomson cross-section, $\sigma_{\rm T}$, the electron mass at rest, $m_{\rm e}$, the speed of light c, the Hydrogen mass fraction, $X = 0.76$, the gas mean molecular weight, μ, and the Hydrogen atom mass, $m_{\rm H}$.

2.5.2 Scaling Relations

In this thesis we study the impact of non-Gaussian models on galaxy cluster scaling relations of temperature, $T_{\rm mw}$, entropy, S, and the Y and Y_X SZ luminosities with the cluster mass, M. Following [180, 181], these scalings can be written as:

$$T_{\rm mw} = A_{\rm TM}\,(M/M_0)^{\alpha_{\rm TM}}\,(1+z)^{\beta_{\rm TM}}\,E(z)^{2/3}\,, \tag{2.45}$$

$$S = A_{\rm SM}\,(M/M_0)^{\alpha_{\rm SM}}\,(1+z)^{\beta_{\rm SM}}\,E(z)^{-2/3}\,, \tag{2.46}$$

$$Y = A_{\rm YT}\,(M/M_0)^{\alpha_{\rm YM}}\,(1+z)^{\beta_{\rm YM}}\,E(z)^{2/3}\,, \tag{2.47}$$

$$Y_X = A_{\rm YxM}\,(M/M_0)^{\alpha_{\rm YxM}}\,(1+z)^{\beta_{\rm YxM}}\,E(z)^{2/3}\,, \tag{2.48}$$

where M_0 was set equal to $10^{14}h^{-1}{\rm M}_\odot$ and all cluster properties are evaluated within R_{200} (see eq. (2.37)). In this way, the redshift evolution of each scaling is modelled by a power of the $E(z)$ function, giving the predicted evolution extrapolated from the self-similar model [170, 171], times a power-law of $(1+z)$ accounting for departures to self-similar evolution. The quantities, A, α, and β, are therefore the scalings normalization at $z = 0$; the mass power-law index; and the index of

the redshift power-law giving the deviation to self-similar evolution, respectively. Whenever $\beta = 0$ the redshift evolution of the scalings is said to be self-similar. Under the assumptions in [170], the self-similar power-law indexes of the mass are $\alpha_{\rm TM} = \alpha_{\rm SM} = 2/3$, and $\alpha_{\rm YM} = \alpha_{\rm YxM} = 5/3$.

To determine A, α, and β for each scaling we use the method described in [178, 181]. This involves re-writting eqs (2.45)–(2.48) in a logaritmic, concise, form,

$$\log(y\, f(z)) = \log(y_0(z)) + \alpha\, \log(x/x_0)\,, \tag{2.49}$$

$$\log(y_0(z)) = \log(A) + \beta\, \log(1+z)\,, \tag{2.50}$$

where y and x are cluster properties, and $f(z)$ is some fixed power of the cosmological factor $E(z)$. The method starts with a fit of the cluster populations at each redshift with eq. (2.49) and minimizing the χ^2 between the the data and the model. If the logarithmic slope α does not change (i.e. shows no systematic variations) with z, the fitting procedure is then repeated with α set to its value at redshift zero, $\alpha(z=0)$, and the scaling normalisation factors $y_0(z)$ are stored. In this way we avoid unwanted correlations between α and the normalizations $y_0(z)$. At this step we also store the r.m.s. dispersion of the fits at each redshift,

$$\sigma_{\log y'} = \sqrt{\frac{1}{N}\sum_i (\log(y'_i/y'))^2}\,, \tag{2.51}$$

where $y' = yf$ (see eq. (2.49)) and y'_i are individual data points. To determine the parameters A and β, we fit eq. (2.50) to the stored values of $\log(y_0(z))$ as a function of $\log(1+z)$, again minimizing the χ^2 between the values $\log(y_0(z))$ and the prediction of eq.(2.50) at each value of $\log(1+z)$. Since cluster abundances drop rapidly with z (see table 2.2), we limited the present cluster scaling analysis to the redshift range $0 \leq z \leq 1$, so that the fitting procedure is carried out with a reasonable number of clusters for all realization runs. We have also checked that the application of this fitting procedure to individual realization catalogues and to single catalogues that combine clusters from realizations runs of each model lead to equivalent results for the derived scalings. We therefore use the latter catalogues to display fitting values and figures, from this point onwards.

2.5.3 Scaling Relations at z=0

In this section we discuss the scalings, eqs (2.45)–(2.48), obtained at redshift zero, from our suite of N-body/hydrodynamic simulations runs with non-Gaussian initial conditions.

Figure 2.8 shows the galaxy cluster distributions for the scalings: $T_{\rm mw} - M$ (top left), $S - M$ (top right), $Y - M$ (bottom left) and $Y_X - M$ (bottom right), with quantities computed within R_{200}. In each panel, models are labeled according to their values of f_{NL}: -500 blue squares, -300 red diamonds, -100 red pentagons, 0 black filled circles, 100 magenta asterisks, 300 cyan pluses and 500 yellow dots. To improve clarity, we only display 500 clusters for each model, randomly drawn from the combined realizations catalogues with a weighting procedure that guarantees that the most massive and rare objects are displayed[1].

[1] The weight attributed to each cluster is proportional to its mass.

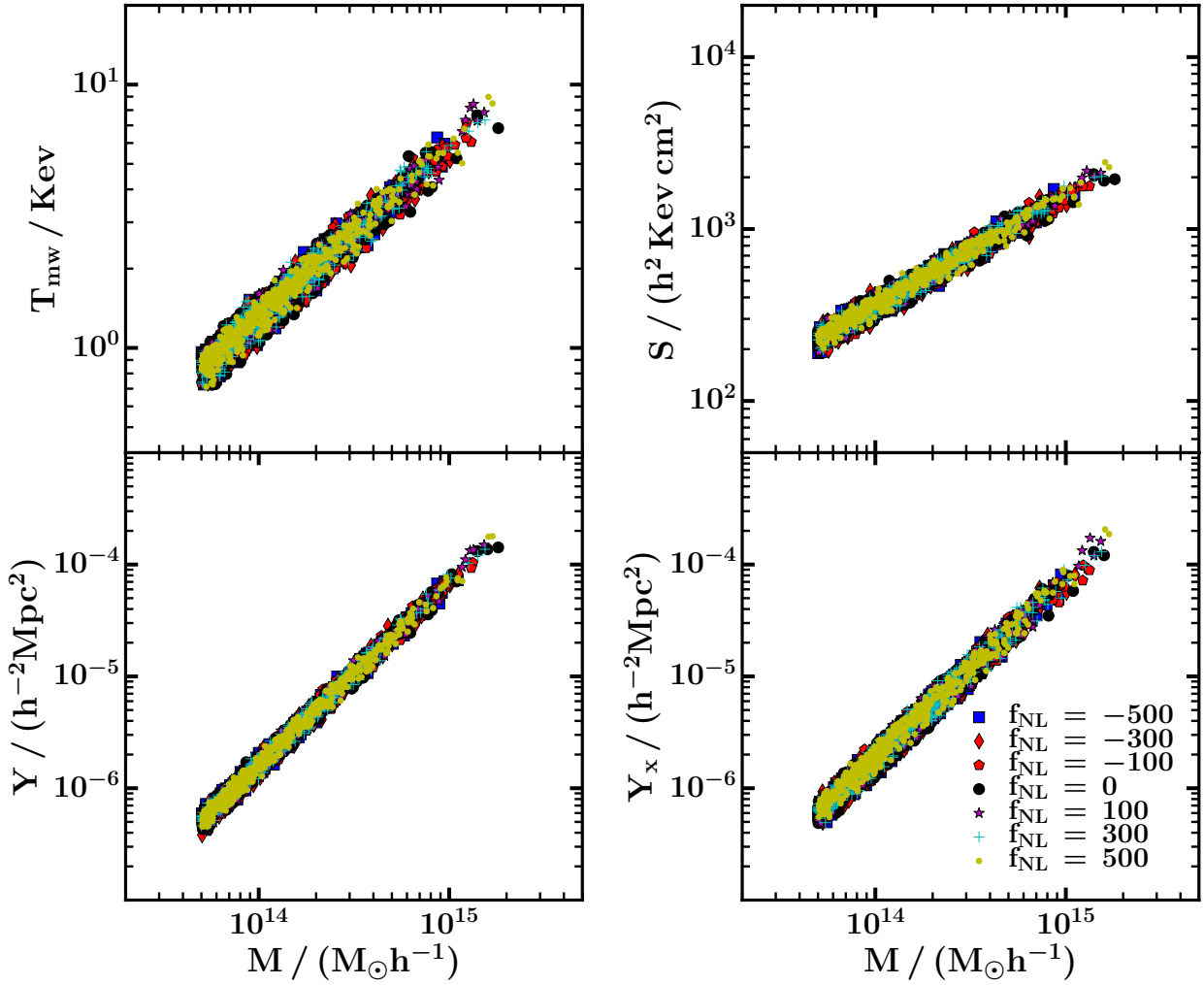


Figure 2.8: Cluster scalings at redshift zero for the $T_{\rm mw} - M$ (top left panel), $S - M$ (top right panel), $Y - M$ (bottom left panel) and $Y_X - M$ (bottom right panel), for values of f_{NL} ranging from -500 to 500 with increments of 200 and a guassian model, $f_{NL} = 0$. The displayed quantities were computed within R_{200}. For clarity, we only plotted 500 clusters randomly selected from the catalogues for each model.

A common trend in all panels is that $T_{\rm mw}$, S, Y and Y_X are properties tightly related to the total cluster mass. This confirms expectations, because temperature is weighted by mass (not by X-ray emission) and entropy is computed using $T_{\rm mw}$ which is a better proxy than T_X for the thermodynamic temperature. On the other hand, the cluster integrated SZ signal is a measure of the total thermal energy of the object, which is known to be more dependent on the cluster total gravitational mass and gas mass fraction than on the details of gas physical effects acting inside R_{200}. The $Y_X - M$ relation displays larger dispersions than the $Y - M$ scaling, because the former is computed using the X-ray emission-weighted temperature, T_X, which is more sensitive to internal gas physical effects than $T_{\rm mw}$.

Table 2.3 presents the best fit parameters, $\alpha(z = 0)$ and log A, and fit dispersions, $\sigma_{\log y'}\,[z = 0]$, of our cluster scaling relations at redshift zero (see Section 2.5.2). In general, all scalings show very

Table 2.3: Best fit values of the parameters α, $\log A$ and β as well as their respective 1σ errors. These values are valid within the redshift range $0 \leq z \leq 1$.

	Model NG-500	Model NG-300	Model NG-100	Model G	Model NG100	Model NG300	Model NG500
$T_{\rm mw} - M$							
$\alpha\,(z=0)$	0.634 ± 0.005	0.638 ± 0.005	0.642 ± 0.005	0.631 ± 0.005	0.644 ± 0.005	0.639 ± 0.005	0.644 ± 0.004
$log\,(A)$	0.116 ± 0.001	0.118 ± 0.001	0.119 ± 0.001	0.120 ± 0.001	0.121 ± 0.001	0.123 ± 0.001	0.124 ± 0.001
β	-0.099 ± 0.003	-0.092 ± 0.003	-0.086 ± 0.005	-0.086 ± 0.004	-0.087 ± 0.004	-0.092 ± 0.003	-0.086 ± 0.004
$\sigma_{\log y'}\,[z=0]$	0.035	0.037	0.036	0.037	0.035	0.036	0.037
$\sigma_{\log y'}\,[z=1]$	0.035	0.036	0.036	0.037	0.038	0.035	0.037
$S - M$							
$\alpha\,(z=0)$	0.629 ± 0.004	0.631 ± 0.004	0.638 ± 0.004	0.630 ± 0.004	0.641 ± 0.004	0.635 ± 0.004	0.643 ± 0.004
$log\,(A)$	2.551 ± 0.001	2.553 ± 0.001	2.555 ± 0.001	2.555 ± 0.001	2.557 ± 0.001	2.559 ± 0.001	2.561 ± 0.001
β	-0.125 ± 0.002	-0.122 ± 0.002	-0.117 ± 0.004	-0.115 ± 0.003	-0.114 ± 0.003	-0.120 ± 0.002	-0.116 ± 0.003
$\sigma_{\log y'}\,[z=0]$	0.031	0.033	0.032	0.033	0.031	0.031	0.032
$\sigma_{\log y'}\,[z=1]$	0.031	0.032	0.031	0.032	0.033	0.031	0.033
$Y - M$							
$\alpha\,(z=0)$	1.640 ± 0.006	1.648 ± 0.006	1.650 ± 0.006	1.632 ± 0.006	1.650 ± 0.006	1.642 ± 0.006	1.644 ± 0.006
$log\,(A)$	-5.793 ± 0.001	-5.792 ± 0.001	-5.791 ± 0.001	-5.790 ± 0.001	-5.789 ± 0.001	-5.787 ± 0.001	-5.787 ± 0.001
β	-0.060 ± 0.005	-0.048 ± 0.005	-0.042 ± 0.007	-0.045 ± 0.005	-0.044 ± 0.006	-0.050 ± 0.004H	-0.042 ± 0.005
$\sigma_{\log y'}\,[z=0]$	0.045	0.047	0.046	0.046	0.046	0.046	0.047
$\sigma_{\log y'}\,[z=1]$	0.045	0.046	0.046	0.046	0.047	0.045	0.047
$Y_X - M$							
$\alpha\,(z=0)$	1.603 ± 0.008	1.607 ± 0.008	1.608 ± 0.008	1.592 ± 0.008	1.616 ± 0.008	1.609 ± 0.008	1.613 ± 0.007
$log\,(A)$	-5.731 ± 0.001	-5.727 ± 0.001	-5.725 ± 0.001	-5.723 ± 0.001	-5.722 ± 0.001	-5.718 ± 0.001	-5.715 ± 0.001
β	-0.127 ± 0.007	-0.118 ± 0.007	-0.107 ± 0.01	-0.111 ± 0.008	-0.103 ± 0.009	-0.112 ± 0.006	-0.108 ± 0.008
$\sigma_{\log y'}\,[z=0]$	0.060	0.062	0.061	0.062	0.061	0.062	0.062
$\sigma_{\log y'}\,[z=1]$	0.066	0.068	0.069	0.067	0.071	0.068	0.071

similar slopes for the various models. Low f_{NL} models seem to have slightly smaller slopes but variations are consistent within one to two 1σ errors giving the statistical uncertainties of the fits.The results for $\log A$ in table 2.3, indicate that the normalization of the scalings at $z = 0$ has a mild but systematic increase with f_{NL}. This is impossible to visualize in the plots of each scaling due to the intrinsic dispersions of the fits. Finally all $z = 0$ scalings show fit dispersions $\sigma_{\log y'}\,[z = 0]$ which are independent of the level of primordial non-Gaussianities. According to table 2.3 the intrinsic dispersion of the $Y_X - M$ scaling is about 1.8 times larger than the dispersion of the $Y - M$ scaling at $z = 0$.

2.5.4 Evolution of Scaling Relations

To study the evolution of the scaling laws we applied the method described in subsection 2.5.2 to the full set of cluster catalogues in our simulations. As mentioned earlier, we carried out the analysis in two ways. One applies the method to catalogues from individual realization runs, from which averaged fitting parameters were inferred for each model. A second approach consisted in combining individual realization catalogues at each redshift and then applying the fitting procedure to the resulting combined catalogues to obtain the scaling parameters. We verified that both approaches lead to equivalent scaling parameters within the defined range of redshifts, $0 \leq z \leq 1$. The results presented here are from the second approach, which somewhat simplifies the presen-

tation of results and the legibility of plots.

The main result of this subsection is the set of plots presented in figure 2.9. These show the evolution of the fitting parameters in (eq. 2.49), the power-law index α and the normalization $\log(y_0(z))$, for all scalings and models considered in this section. The figure is divided in four plots, one for each scaling (top left: $T_{\rm mw} - M$; top right: $Y - M$; bottom left: $S - M$; bottom right: $Y_X - M$). Each plot contains two panels displaying the evolution of α (top panel) and $\log(y_0(z))$ (bottom panel) with z. Models are labeled in the same way as in figure 2.8 and bars in the data points are 1σ bootstrap resampling errors.

A first conclusion from figure 2.9 is that the power-law index, α shows no systematic variation with z for all scalings. In general, data points and errors appear scattered around the redshift zero value, $\alpha(z = 0)$, for each f_{NL} model in all scalings. We note that although our simulations include only adiabatic gas physics all α points (including those from the Gaussian Λ model) are, in general, below the self-similar predictions: $\alpha_{\rm TM} = \alpha_{\rm SM} = 2/3$, and $\alpha_{\rm YM} = \alpha_{\rm YxM} = 5/3$. These predictions assume hypothesis such as hydrostatic equilibrium and spherical symmetry in clusters (as well as a critical density cosmology, [170]) which are only approximations to the true state of clusters in simulations [171]. Deviations from self-similar values are small but in most cases larger than the statistical errors. The larger deviations are found for the $Y_X - M$ scaling, which presents systematically lower α than the $Y - M$ scaling. This is because the SZ signal is proportional to the product of the cluster gas temperature by mass ($Y_{\rm SZ} \propto TM$) and the temperature scales in our simulations as $T_{\rm mw} \propto M^{0.64}$ and $T_X \propto M^{0.60}$ (these values are good approximations for all models). Here we will not display further results for the $T_X - M$ scaling, which has an evolution for the Gaussian model consistent with the results in figure 2 for the $w = -1$ simulations in [181] (their simulations have a smaller boxsize but the same gas physics and similar cosmology to our G model runs).

The scaling-law normalizations, $\log(y_0(z))$, in figure 2.9 denote clear trends with redshift and f_{NL}. The decrease of $\log(y_0(z))$ with z puts in evidence that all scalings tend to deviate from self-similar evolution, in a way that clusters of a given mass have lower temperatures, entropies and $Y_{\rm SZ}$ signals at higher z than what would be expected assuming self-similar evolution. The panels show that this negative (with respect to self-similar) evolution follows, in general, linear trends with z that can be fit with eq. (2.50) using the method described in Section 2.5.2. Table 2.3, lists the normalization constant, A and the power-law index β modelling the redshift dependence of $\log(y_0(z))$ obtained in this way for all scalings. These numbers confirm negative β slopes with mild (but statistically significant) deviations from the self-similar expectation $\beta = 0$. The dependence of the $\log(y_0(z))$ normalization with f_{NL} is also evident from figure 2.9. For each scaling, models with higher f_{NL} tend to show larger normalizations at all redshifts. This can be understood in light of the findings in N-body simulations [160] and analytical modelling using excursion set theory [172] that cluster haloes in non-Gaussian models have increased/decreased core densities for positive/negative f_{NL}. As a consequence cluster gas properties such as temperature, entropy and the$Y_{\rm SZ}$ signal are expected to follow this trend, leading to scaling normalizations that increase with f_{NL}.

An interesting aspect to address with cluster simulations is to investigate the evolution of the

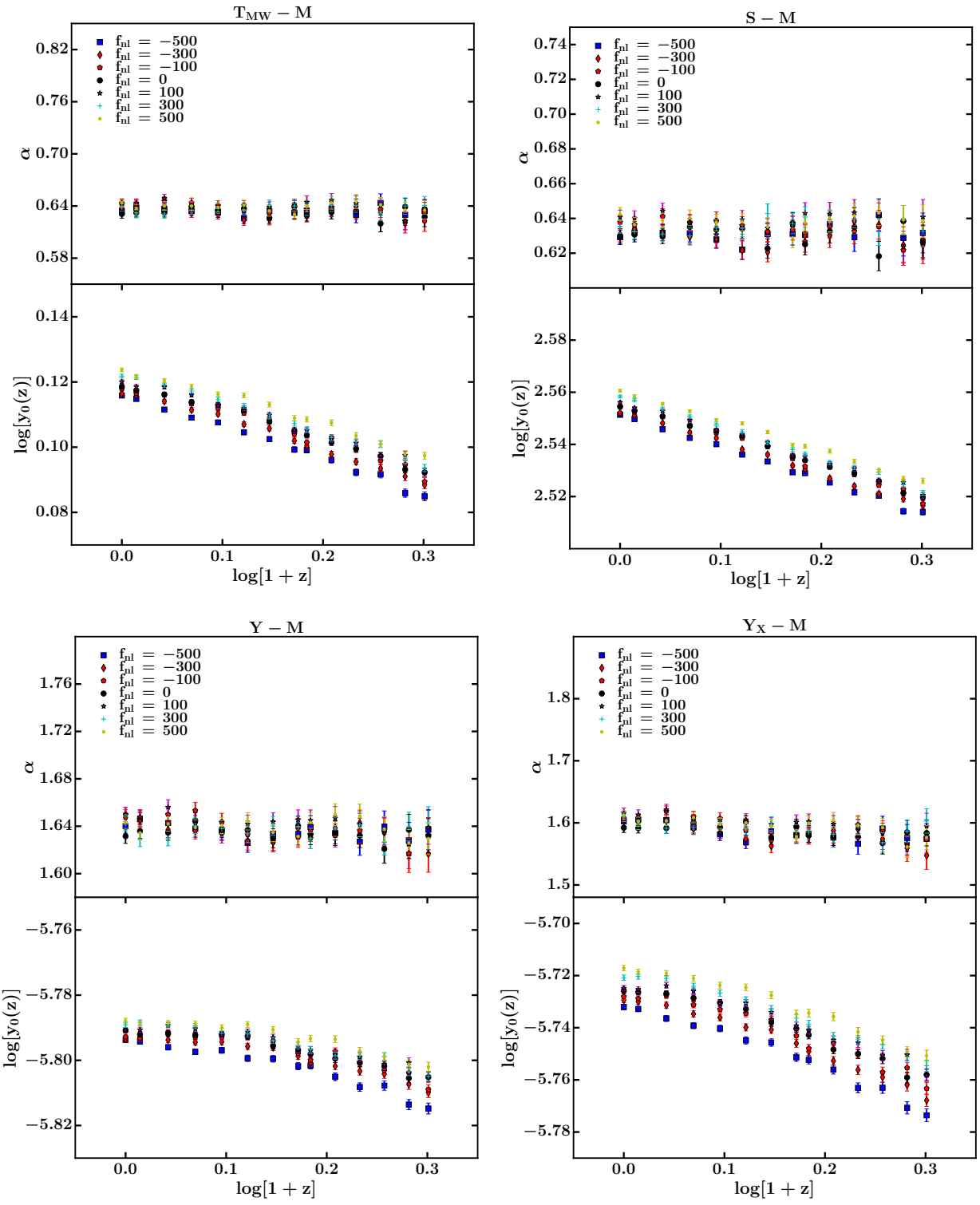


Figure 2.9: The evolution of the slope, α, normalization, $\log_{10}[1+z]$ and respective 1σ error bars, with redshift for $T_{\rm mw} - M$ (top left panel), $Y - M$, (top right panel), $S - M$ (bottom left panel) and $L_X - M$ (bottom right panel), for different values of f_{NL} ranging from -500 to 500 with increments of 100.

intrinsic scatter of scaling laws with redshift. In our simulations we find that cluster scaling laws involving mass-weighted quantities (i.e. the $T_{\rm mw} - M$, $S - M$ and $Y - M$ scalings) show no significant evolution of fit dispersions, $\sigma_{\log y'}$, with redshift. The quantities $\sigma_{\log y'}\,[z = 0]$ and $\sigma_{\log y'}\,[z = 1]$ in table 2.3 give the fit dispersions in our models at $z = 0$ and $z = 1$, respectively. For the $Y_{\rm X} - M$ scaling our simulations indicate an increase of the fit dispersions with z. This effect is independent of f_{NL} and is related to the fact that Y_X depends on T_X, which in turn is a function the evolution of gas X-ray emission with redshift.

2.5.5 Dependence on f_{NL}

With the results from table 2.3 we constructed plots in figure 2.10 that put in evidence the impact of non-Gaussian initial conditions on the four galaxy cluster scaling-laws investigated in this thesis. The panels in each plot give the best fit values for the mass power-law index α (top panel), the scaling normalization $\log A$ (centre panel), and the $(1 + z)$ power-law index β (bottom panel) as a function of f_{NL}. Data in black are the results from table 2.3. To test the robustness of the results with respect to a different choice of M_{lim} we repeated the analysis in the previous sections imposing a higher minimum mass limit, $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$, to our simulated catalogues. This analysis leads to the data displayed in red. For both colour-coded data sets, bars indicate bootstrap errors, lines are straight-line fits to the data points, and shaded areas represent the 95% confidence levels preferred by the data.

These plots indicate that the mass power-law index α remains approximately unchanged with f_{NL}. Variations are as small as 1.9%, 1.3%, 0.6% and 0.1% for the scaling $S - M$, $T_{\rm mw} - M$, $Y_X - M$ and $Y - M$, respectively. When less massive clusters and groups are excluded from the analysis (see data points from the $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$ catalogues), the dependence of α on f_{NL} is even weaker for $T_{\rm mw} - M$, $S - M$ and $Y - M$ scalings, with variations of about 0.9%, 0.3% and 0.3% respectively; while for $Y_X - M$ scalings the dependence is slightly stronger but not larger than 1%. This means that the α variations with f_{NL} in our $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$ catalogues are always below the one percent level for all scalings.

The scaling laws normalization parameter A is slightly more sensitive to non-Gaussianities. Within the displayed range of f_{NL}, the normalization parameters A change by about 3.8% for $Y_X - M$, 2.1% for $S - M$, 1.9% for $T_{\rm mw} - M$ and 1.6% for $Y - M$ scalings. Similar variations are found for the results obtained with $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$ catalogues.

The impact of non-Gaussian initial conditions is stronger for the redshift power-law index, β, that measures the departures from self-similar evolution of the scalings. The variations of β within the displayed range of f_{NL} are about 20%, 13%, 11.4% and 6.5% for the scalings $Y - M$, $Y_X - M$, $T_{\rm mw} - M$ and $S - M$, respectively. When the less massive objects are excluded from the analysis (catalogues with $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$), the $T_{\rm mw} - M$ and $S - M$ scalings show weaker variations with f_{NL}. The SZ scalings show slightly larger percentage variations but systematically lower β when compared with the results from the $M_{lim} = 5 \times 10^{13} h^{-1} M_{\odot}$ catalogues.

The effect of non-Gaussian initial conditions on these cluster scalings is consistent with the view that positive/negative f_{NL} tend to increase/decrease cluster concentrations [160]. Clusters with

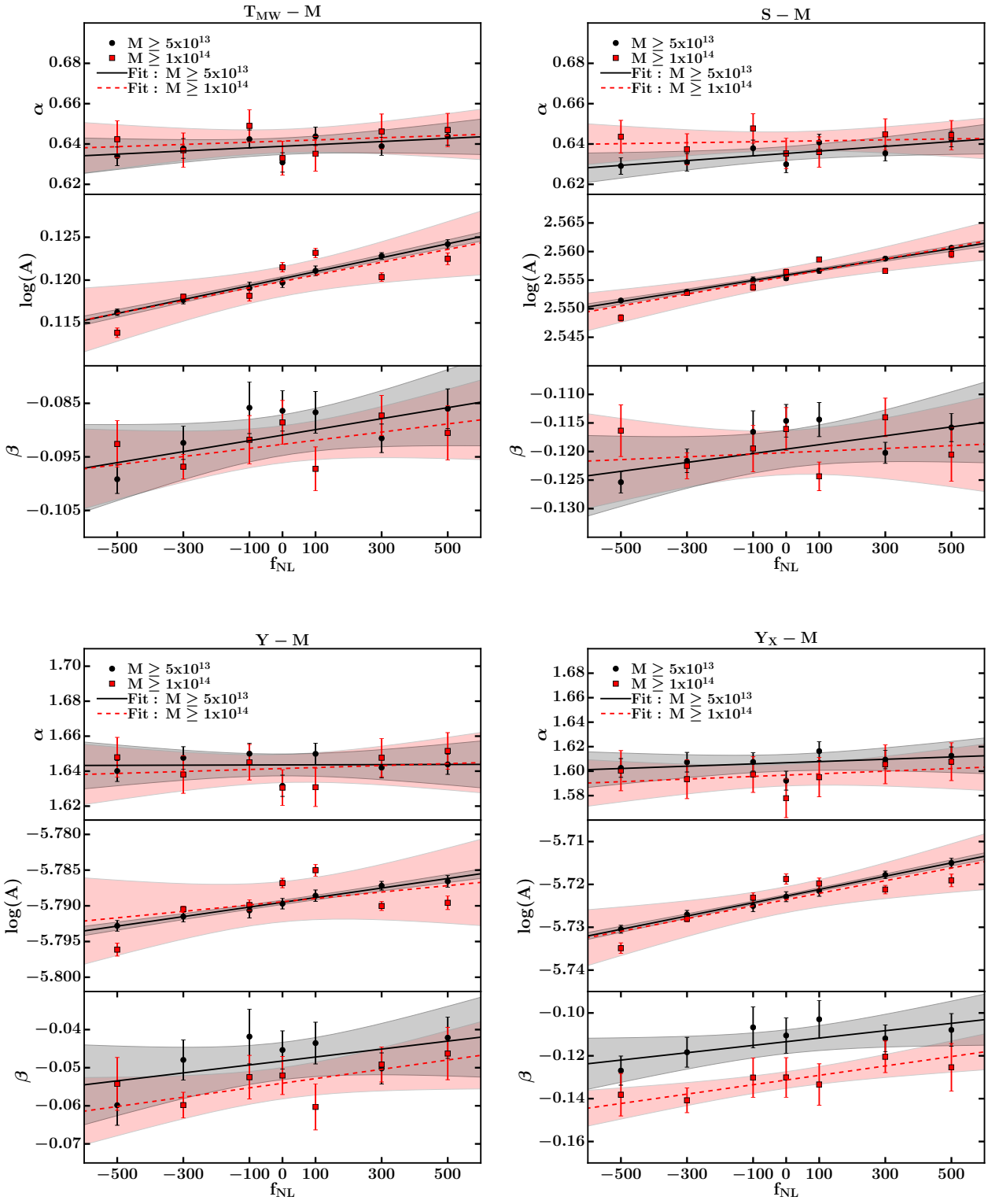


Figure 2.10: The dependence of the power-law index α of mass, normalization parameter $\log_{10}(A)$, and power-law index of redshift β and their respective 1σ error bars, as a function of f_{NL} for $T_{\rm mw} - M$ (top left panel), $Y - M$, (top right panel), $S - M$ (bottom left panel) and $Y_X - M$ (bottom right panel). Black solid line and black shaded area corresponds to the linear fit and 95% C.L. confidence interval for a mass cut of $5 \times 10^{13}\, M_{\odot} h^{-1}$, while the dashed red line and red shaded area corresponds to the linear fit and 95% C.L. confidence interval for a mass cut of $1 \times 10^{14}\, M_{\odot} h^{-1}$.

higher concentrations tend to have higher gas densities and temperatures (and therefore higher entropy and the $Y_{\rm SZ}$ signal) at the their inner regions. According to our findings, this influences the normalization A and the evolution β parameters of the cluster scaling laws. We note that, although f_{NL} has a significant impact on β, these departures from self-similar evolution are in general small for all scalings. According to [160] the effect of non-Gaussianity on cluster concentrations increases slightly with mass. This effect appears not to have a too strong impact on the cluster fitting parameters when we change the minimum mass limit of our catalogues to $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$. The exception may be the β parameters in the $Y_{\rm SZ}$-mass scalings, which show a slight increase when low-mass clusters and groups are excluded from the catalogues. This tendency is however reversed in the case of the $T_{\rm mw} - M$ and $S - M$ scalings. We note, however, that the effect of cluster concentrations is in competition with other effects such as the increase of scatter due to a reduction of the total number of clusters in the fitting procedure when the minimum mass limit of the catalogues is increased to $M_{lim} = 1 \times 10^{14} h^{-1} M_{\odot}$.

2.6 Conclusions

In this chapter we have investigated whether the presence of primordial non-Gaussianities has an impact on the estimation of the effective dark energy equation of state, when one uses the abundance of galaxy clusters as a tool to probe different cosmological scenarios. We computed the effective dark energy equation of state, w_{eff} per redshift bin, assuming Gaussian initial conditions, that is capable of reproducing the galaxy cluster counts expected in several non-Gaussian models, thus constructing a correspondence $f_{NL} \mapsto w_{eff}$ for each redshift bin. The most important result of this work is the discovery of a redshift interval where no value for the effective dark energy equation of state is capable of reproducing the non-Gaussian cluster abundance for models with $f_{NL} > 0$, which is the result of there being a w_{eff} that maximizes the cluster number density at each redshift, while for models with $f_{NL} < 0$ a discontinuous w_{eff} is obtained. The appearance of such type of features may thus constitute a new diagnostic of the presence of primordial non-Gaussianities. Although, even under the ideal situation where only statistical uncertainties are present, their detection is only possible for values of f_{NL} not too far away from those permitted by analysis of the WMAP 7-year data [32], it should be remembered that f_{NL} may not be invariant with scale. In particular, there isn't any particular reason why f_{NL} could not increase as the scale diminishes (see e.g. [46, 182]).

We have also computed the dependence, as a function of f_{NL}, of the biases that arise in the estimation of the cosmological parameters Ω_{de0}, w and σ_8, when it is (wrongly) assumed, for the purpose of the statistical analysis, that the density field is Gaussian distributed. We have found that such biases are quite small for the first two parameters, but significant in the case of σ_8, in particular in the face of the high statistical accuracy with which this parameter is expected to be determined in the near-future (e.g. [38]). If f_{NL} is assumed to be scale-independent, then the results obtained by the Planck team [34], impose severe constraints on the amount of non-Gaussianity at cluster scales ($f_{NL} = 2.7 \pm 5.8$ for the Local parametrization) making it safe to neglect primordial non-

Gaussianities in the determination of cosmological parameters using the galaxy cluster abundance. However, f_{NL} could be a function of scale (see [46, 182] and references therein) , in which case the effects of non-Gaussianity on the cluster abundance may need to be taken into account in the determination of σ_8. In fact, eq. (2.35) suggests that a significantly negative value for f_{NL} at cluster scales, of the order of $f_{NL} \backsimeq -240$, would reconcile the constraint on σ_8 obtained using the cluster abundance inferred from the Planck Sunyaev-Zel'dovich Catalogue ($\sigma_8 = 0.77 \pm 0.02$, [183]) with the one obtained from the properties of the primordial CMB temperature anisotropies ($\sigma_8 = 0.829 \pm 0.012$, [14]).

Still within the thematic of primordial non-Gaussianities, we present galaxy cluster scaling relations from hydrodynamic/N-body simulations of large-scale structure, featuring adiabatic gas physics and non-Gaussian initial conditions for the mater density fluctuations. We investigated five non-Gaussian models with Local f_{NL} parametrizations ranging from -500 to 500 and a $f_{NL} = 0$ Gaussian model, with a flat ΛCDM cosmology. We did a total of 35 simulation runs and generated catalogues with cluster masses larger than $M_{lim} = 5 \times 10^{13}\, h^{-1} M_{\odot}$ to study scaling relations involving mass-weighted temperature, $T_{\rm mw}$, entropy, S, integrated SZ signals, Y and Y_X, with mass, M (see Eqs (2.45)–(2.48)).

The main conclusions of this study are:

- Non-Gaussian initial conditions have a mild but significant impact on the normalization of the cluster scalings, $y_0(z)$, and almost no impact on the power-law index, α, of the mass dependence.

- The normalizations $y_0(z)$ are affected by non-Gaussianities through changes in the amplitude parameter A, giving the normalization of the scalings at $z = 0$, and through variations in their non self-similar evolutions, parametrized by power-laws of $(1 + z)$ with indexes β.

- The redshift zero normalizations, A, show only slow increases with f_{NL}, of the order of $1.6\% - 3.8\%$ in the range $-500 \leq f_{NL} \leq 500$, for the various scalings.

- Non-Gaussianities have a stronger impact on the redshift evolution of the normalizations. Our β parameters increase with f_{NL} by a maximum of 20% for the $Y - M$ scaling and a minimum of 6.5% for the $S - M$ scaling within $-500 \leq f_{NL} \leq 500$. In all cases the β parameters, that measure departures from self-similar evolution, are found to be close to the expected self-similar evolution of each scaling.

- Increasing the minimum mass limit of our catalogues to $M_{lim} = 1 \times 10^{14}\, h^{-1} M_{\odot}$, we find similar dependences for A and α with f_{NL}. The dependence of β with f_{NL} becomes stronger for $Y_X - M$ and $Y - M$ and less prominent for the $T_{\rm mw} - M$ and $S - M$ scalings.

These results are in line with the predictions that f_{NL} changes the internal structure of cluster profiles, as a result of an increase/decrease of cluster concentrations for positive/negative f_{NL} [160, 172]. The impact on cluster scaling relations is mostly due to changes in the evolution of their normalizations. Our results show that this impact is small for models with low f_{NL}. However, for larger values of f_{NL}, the effect of PNG on the evolution of cluster scalings can be as important

as the effect of non-gravitational gas physics (see eg [180, 184]) or the effect of dark energy (see, eg [181]) in clusters scaling relations. This means that it is safe to neglect the effects of PNG on the investigated clusters scalings if the present observational constraints from Planck [34, 35] on a scale invariant f_{NL} are valid at galaxy cluster scales. This may no longer be true if f_{NL} is in fact a scale dependent parameter (see e.g. [46, 182]). In this case, $|f_{NL}|$ may have a larger amplitude at clusters scales and therefore our results show that galaxy cluster scalings are sensitive to primordial non-Gaussianities and should be taken into consideration when assessing the constraining power of cluster surveys or when using future galaxy cluster data to infer cosmological parameters.

Chapter 3

COSMOABC: Likelihood-free inference via Population Monte Carlo Approximate Bayesian Computation

The precision era of cosmology marks the transition from a data-deprived field to a data-driven science on which statistical methods play a central role. The ever-increasing data deluge must be tackled with new and innovative statistical methods in order to improve our understanding of the key ingredients driving our Universe (see e.g. [185–188]). Given the continuous inflow of new data, one does not start an analysis from scratch for every new telescope, but is guided by previous knowledge accumulated through experience. A new experiment provides extra information which needs to be incorporated into the larger picture, representing a small update on the previous body of knowledge. Such a learning process is a canonical scenario to be embedded in a Bayesian framework, which allow us to update our degree of belief on a set of model parameters[1] whenever new and independent data are acquired.

A standard Bayesian analysis specifies prior distributions on unknown parameters, defines which parameter values better describe the relationship between the model, the prior and the new data, and then finds the posterior distribution – either analytically or via sampling techniques, as e.g. with Markov Chain Monte Carlo (MCMC) [190]. This analysis requires a proper construction of the likelihood function, which is not always well known or easy to handle. A common solution would be to construct a model for the likelihood (e.g. a Gaussian) followed by MCMC, with the expectation that this hypothesis is not too far from the true. Nonetheless, the challenge of performing parameter inference from an unknown or intractable likelihood function is becoming familiar to the modern astronomer. Recent efforts to overcome observational selection biases in the study of massive [191] and not-so massive [192] stars, to account for windowing effects, errors and/or gaps in time-series of X-ray emission from active galactic nuclei [193, 194] and UV emission from stellar coronae

[1] For the purposes of this work, we will only be interested in the *parameter values* of a given model. However, it is important to stress that in a completely Bayesian approach all the elements and hypotheses forming the model can be considered part of the *prior*. In this sense, with the arrival of new essential information, the Bayesian approach allows for completely redefinition of the model itself [189].

[195] have been reported.

The development of *ad hoc* approaches to this problem within astronomy have proceeded independently from their long history within the field of population genetics. The latter have ultimately been formalized into a rigorous statistical technique known as Approximate Bayesian Computation (ABC). The intuition of ABC dates back to a thought-experiment in [196], where the basic ABC rejection sampler is used to illustrate Bayes Theorem. The authors of [197] employ an acceptance-rejection method in the context of population genetics, while [198] presents the first implementation of a basic ABC algorithm. Only recently the ABC approach has been introduced and applied to astronomical problems [199–202]. This work is part of a larger endeavour and a natural consequence of such initial efforts. Following the philosophy behind the *Cosmostatistics Initiative* (COIN)[2], we present a tool which enables astronomers to easily introduce ABC techniques into their daily research.

The cornerstone of the ABC approach is our capability of performing quick and reliable computer simulations which mimic the observed data in the best possible way (this is called *forward simulation inference*). In this context, our task relies on performing a large number of simulations and quantifying the "distance" between the simulated and observed catalogues. The better a parametrization reproduces the observed data in a simulated context, the closer it is to the "true" model. From this simple reasoning, many alternatives were developed to optimize the parameter space sampling and the definition of distance. One of such examples is the work of [203], who proposes a merger between the standard MCMC algorithm and the ABC rejection sampling. In astronomy, the method was used by [202] to constrain the Milky Way thick disk formation. Going one step further, [204] propose to evolve an initial set of parameter values (or *particle system*) through incremental approximations to the true posterior distribution. The *Population Monte Carlo ABC* (PMC-ABC) method was used to make inferences on rate of morphological transformation of galaxies at high redshift [199] and proved to be efficient in tracking the Hubble parameter evolution from type Ia supernova measurements, despite the contamination from type II supernova [201]. More recently, [205] used ABC to predict weak lensing peak counts and [206] applied a weighted variant of the algorithm to cluster strong lensing cosmology.

The work done throughout this chapter, introduces COSMOABC[3], the first publicly available[4] Python ABC package for astronomy[5]. The package is structured so that the simulation, priors and distance functions are given as input to the main PMC-ABC sampler. In this context, users can easily connect the ABC algorithm to their own simulator and verify the effectiveness of the tool in their own astronomical problems. The package also contains exploratory tools which help defining a meaningful distance function and consequently point to appropriate choices before the sampler itself is initiated.

At first glance, the issues presented in this chapter seems to be disconnected from the main theme of the thesis. However, it is important note that in section 2.4 of the previous chapter, we

[2] http://goo.gl/rQZSAB

[3] https://pypi.python.org/pypi/CosmoABC

[4] Shortly after COSMOABC was released [207] also presented a Python package for forward modelling through PMC-ABC.

[5] For similar tools in the context of biology and genetics, see e.g. [208, 209].

made use of Monte Carlo Markov Chain techniques to explore the posterior in the parameter space. Furthermore, we have assumed that we knew both the mass of the clusters and their location in redshift with infinite precision. When dealing with observations, the latter is no longer true. In this situation, those quantities will have to be estimated in some way from the data and it will be inevitable to treat them within the Bayesian formalism. Moreover, even the assumption that the abundance of galaxy clusters follows a Poisson distribution may not hold, since spatial correlations between clusters have to be taken into account. Hence, as more and more layers of complexity are introduced in our model, the more complex the likelihood will become, thus motivating the search for alternative ways to perform Bayesian cosmological parameter estimation.

The outline of this chapter is as follows. In section 3.1, we give an overview of the Bayesian perspective and the ABC algorithm. In section 3.2 we present the COSMOABC package, through a simple toy model, in order to explain how the algorithm and the package work. Finally in section 3.3 we give a practical application of COSMOABC to a cosmological problem and show how the machinery can be used to define credible intervals over cosmological parameters based on galaxy clusters catalogues. Simulations of the galaxy cluster catalogues for this example were performed using the *Numerical Cosmology* library NUMCOSMO [210] and thus we also exemplify how to link both packages to obtain constrains on cosmological parameters from galaxy cluster number counts.

3.1 Bayesian approaches to parameter inference

Statistical inference on unknown parameters is often a primary goal of a physical experiment design. Although it is possible, and at times even desirable, to encounter some unpredictable behaviour among the outcomes of an experiment or a measurement, in many situations the experimentation aims at establishing constraints over the parameters of a model. In other words, the desire is to use real-world data to check prevailing theories.

In the Bayesian framework, the data are seen as the accessible truth regarding a given physical process and the model as a representation of our understanding of such process. This approach is data-centred and allows us to update the model whenever new information becomes available. In other words, our goal is to determine the probability of a model given the data,

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})}, \tag{3.1}$$

where θ is the vector of model parameters, $\mathcal{D}$ the data set, $p(\boldsymbol{\theta}|\mathcal{D})$ is called the *posterior*, the prior, $p(\boldsymbol{\theta})$, represents our initial expectations towards the model and $p(\mathcal{D})$ is a normalization constant.

In this context, the model parameters are treated as random variables and each individual measurement corresponds to one realization of them. Thus, once our prior is confronted with the data, the outcome is a posterior PDF. Using the posterior distributions we can determine *credible intervals*, which represent our uncertainty about the model parameters[6]. For example, one may

[6]Not to be confused with the frequentist definition *confidence interval*, where the parameter values are considered fixed and therefore, there is no probabilistic interpretation associated to them.

be interested in the most-probable region of values for certain parameters.

3.1.1 Approximate Bayesian Computation

The ABC algorithm uses our ability to simulate the physical process under investigation in order to bypass the necessity of an unknown or computationally too expensive likelihood function. It is based on the following crucial elements:

- a simulator, or forward model,
- prior probability distributions over the input parameters $p(\boldsymbol{\theta})$,
- a distance function, $\rho(\mathcal{D}_1, \mathcal{D}_2)$[7].

As a simple example, consider the following toy model: a given physical process can be probed through a catalogue of P observations, $\mathcal{D} = \{x_i, ..., x_P\}$. Our model states that this process is driven by a random variable, $\mathcal{X}$, following a Gaussian distribution, $\mathcal{X} \sim \mathcal{N}(\mu_0, \sigma_0)$. Thus our goal is to identify credible intervals over μ_0 and σ_0 based on $\mathcal{D}$. Moreover, our prior states that $\mu_0 \in [\mu_-, \mu_+]$ and $\sigma_0 \in [\sigma_-, \sigma_+]$. Hereafter, we will denote the model parameters as $\boldsymbol{\theta} = \{\mu, \sigma\}$.

The main idea behind the ABC algorithm can be summarized in three main steps:

- draw a large number of parameter values, $\boldsymbol{\theta}^i$, from the prior distribution, p,
- for each $\boldsymbol{\theta}^i$ generate a simulation, $\mathcal{D}^i_{\mathrm{S}}$, and calculate the distance between the observed to the simulated catalogues, $\rho^i = \rho(\mathcal{D}, \mathcal{D}^i_{\mathrm{S}})$,
- approximate the posterior probability distribution using the fraction of $\boldsymbol{\theta}^i$'s with smallest associated distances.

The above method has been modified and further developed in the last decade, generating some alternatives to the main algorithm (e.g. [211–215]). One of them is presented below.

3.1.2 Distance

In the toy model described above (see subsection 3.1.1) , we can safely determine the distance, ρ, between the measured catalogue $\mathcal{D}$ and a simulated one $\mathcal{D}_{\mathrm{S}}$ as

$$\rho = \mathrm{abs}\left(\frac{\bar{\mathcal{D}} - \bar{\mathcal{D}_{\mathrm{S}}}}{\bar{\mathcal{D}}}\right) + \mathrm{abs}\left(\frac{\sigma_{\mathcal{D}} - \sigma_{\mathcal{D}_{\mathrm{S}}}}{\sigma_{\mathcal{D}}}\right), \tag{3.2}$$

where $\bar{\mathcal{D}}$ is the mean of all measurements in catalogue $\mathcal{D}$ and $\sigma_{\mathcal{D}}$ is its standard deviation. Eq. 3.2 encloses important properties, which should be present in any ABC distance function: the distance between two identical catalogues is zero and the distance value increases steeply as parameter values get further from the fiducial ones. We emphasise that the choice of the distance function is a crucial step in the design of the ABC algorithm. Thus, its properties must be carefully checked before any ABC implementation is attempted.

[7] Not to mistake with the definition of density in the previous chapters.

3.1.3 Population Monte Carlo ABC

COSMOABC uses the algorithm proposed by [204], where successive steps towards the posterior are achieved by applying an importance (or weighted) sampling in the set of parameter values whose distances satisfy a given initial threshold.

We begin by drawing M values from the prior, called *particles*, $\{\boldsymbol{\theta}^i\}$ with $i \in [1, M]$, such that $M >> N$ (N is the number of samples needed to characterize the prior). Furthermore, since each particle is independently drawn from the prior, the generation of the set of M particles can be speeded up by parallelizing this step and span the work load through different CPU's. For each particle we generate a forward model (simulation) and calculate the distance between synthetic and real catalogues $\rho^i = \rho(\mathcal{D}, \mathcal{D}_S^i)$. From this large set, we keep only the N particles with smallest ρ^i, which constitute the first *particle system* ($\mathcal{S}_{t=0}$) and determine a distance threshold for the next iteration ($\epsilon_{t=1}$) as the 75% quantile of all $\rho \in \mathcal{S}_{t=0}$. In this initial step, we associate to each particle the same weight, $W_{t=0}^j = 1.0/N$, for $j \in [1, N]$.

In subsequent iterations, $t > 0$, we perform an *importance sampling* from $\mathcal{S}_{t-1}$: a popular technique where one can draw from a proposal distribution and re-weight the particle system so it targets the desired posterior distribution.

The parameter vector resulting from this importance sampling, $\boldsymbol{\theta}_{\text{try}}$, is used to simulate a catalogue and calculate its distance to the observed data, ρ_{try}. The parameter $\boldsymbol{\theta}_{\text{try}}$ is stored if $\rho_{\text{try}} \leq \epsilon_t$. This process is repeated until a new set of N parameter values satisfying the distance threshold is completed. For the new particle system, the weights are calculated as

$$W_t^j = \frac{p(\boldsymbol{\theta}_t^j)}{\sum_{i=1}^{N} W_{t-1}^i \mathcal{N}(\boldsymbol{\theta}_t^j; \boldsymbol{\theta}_{t-1}^i, C_{t-1})}, \tag{3.3}$$

where W_t^j denotes the weight associated to the $j-th$ particle in particle system t, $p(\boldsymbol{\theta}_t^j)$ corresponds to the prior probability distribution calculated at $\boldsymbol{\theta}_t^j$, W_{t-1}^i is the weight of the $i-th$ particle in particle system $t-1$ and $\mathcal{N}(\boldsymbol{\theta}^j; \boldsymbol{\theta}_{t-1}^j, C_{t-1})$ represents a Gaussian PDF[8] centred in $\boldsymbol{\theta}_{t-1}^i$, with covariance matrix built from $\mathcal{S}_{t-1}$ and calculated at $\boldsymbol{\theta}^j$.

Once the new weights are determined, we start the construction of a new particle system and the algorithm is repeated until convergence. As pointed out by [216], this is achieved when the ABC posterior no longer changes substantially with subsequent iterations. Here we consider that the system converged when the number of draws necessary to construct a particle system is much larger than N (see Algorithm 1, Table 3.1 and section 3.2.2). Each iteration brings us closer to the "true" PDF bypassing the need of a full likelihood calculation. Moreover, as the calculation of one particle is independent from the others within each iteration, the algorithm itself is more easily parallelizable than a standard MCMC.

[8] In general, the Gaussian PDF works well, but can be replaced with a different distribution if the parameter space has special restrictions, e.g. only takes integer values.

Data: $\mathcal{D} \longrightarrow$ observed catalogue.
Result: ABC-posteriors distributions over the model parameters.
$t \longleftarrow 0$
$K \longleftarrow M$
for $J = 1, \ldots, M$ **do**
  Draw θ, from the prior, $p(\theta)$.
  Use θ to generate $\mathcal{D}_{\mathrm{S}}$.
  Calculate distance, $\rho = \rho(\mathcal{D}_{\mathrm{S}}, \mathcal{D})$.
  Store parameter and distance values, $\mathcal{S}_{\mathrm{ini}} \leftarrow \{\theta, \rho\}$
end
Sort elements in $\mathcal{S}_{\mathrm{ini}}$ by $|\rho|$.
Keep only the N parameter values with lower distance in $\mathcal{S}_{t=0}$.
$C_{t=0} \longleftarrow$ covariance matrix from $\mathcal{S}_{t=0}$
for $L = 1, \ldots, N$ **do**
  $W_1^L \longleftarrow 1/N$
end
while $N/K > \Delta$ **do**
  $K \longleftarrow 0.$
  $t \longleftarrow t + 1.$
  $\mathcal{S}_t \longleftarrow []$
  $\epsilon_t \longleftarrow 75^{th}$-quantile of distances in $\mathcal{S}_{t-1}$.
  while len($\mathcal{S}_t$) $<$ N **do**
    $K \longleftarrow K + 1$
    Draw θ_0 from $\mathcal{S}_{t-1}$ with weights $\bar{W}_{t-1}$.
    Draw θ, from $\mathcal{N}(\theta_0, C_{t-1})$.
    Use θ to generate $\mathcal{D}_{\mathrm{S}}$.
    Calculate distance, $\rho = \rho(\mathcal{D}_{\mathrm{S}}, \mathcal{D})$
    if $\rho \leq \epsilon_t$ **then**
      $\mathcal{S}_t \longleftarrow \{\theta, \rho, K\}$
      $K \longleftarrow 0$
    end
  end
  for $J = 1, \ldots, N$ **do**
    $\tilde{W}_t^J \longleftarrow$ equation (3.3).
  end
  $W_t \longleftarrow$ normalized weights.
  $C_t \longleftarrow$ weighted covariance matrix from $\{\mathcal{S}_t, W_t\}$.
end

Algorithm 1: PMC-ABC algorithm implemented in COSMOABC.

Table 3.1: Glossary for algorithm 1.

Parameter	Description
$\mathcal{D}$	Observed data set
$\mathcal{D}_S$	Simulated catalogue
M	Number of draws for the first iteration
$\mathcal{S}$	Particle system
N	Number of particles in $\mathcal{S}$
t	Time-step (iteration) index
K	Number of draws index
W	Importance weights
ϵ	Distance threshold
Δ	Convergence criterion
θ	Vector of model parameters
$p(\cdot)$	Prior distribution
$\rho(\cdot,\cdot)$	Distance function
$\mathcal{N}(\bullet;\theta,C)$	Gaussian PDF at $\bullet$ with $\mu=\theta$, cov$=C$

3.2 COSMOABC

In COSMOABC, our toy model can be represented by a simulation function,

```
from scipy.stats import norm
import numpy as np

def my_simulation(v):
    """ Toy model simulator """

    dist = norm(loc=v['mean'],
                scale=v['std'])
    l1 = dist.rvs(size=v['n'])

    return np.atleast_2d(l1).T
```

where v is a dictionary of input parameters whose keywords mean and std determine the mean and standard deviation of the underlying Gaussian distribution, respectively, and n denotes the total number of objects in the catalogue. Analogously, a flat prior would be written as[9]

```
from scipy.stats import uniform

def my_prior(par, func=False):
    """Flat prior"""

    gap = par['pmax'] - par['pmin']
```

[9]The func argument is needed so we can retrieve a realization and the probability distribution itself. This is used by COSMOABC in the calculation of the weights.

```
    pdf = uniform(loc=par['pmin'],
                  scale=gap)
    if func == False:
        draw = pdf.rvs()
        return draw
    else:
        return pdf
```

with par as a dictionary of input parameters and the keys pmin and pmax determining the boundaries of the distribution.

The distance function should be written as

```
import numpy as np

def my_distance(d2, p):
    """Distance function."""

    dmean = np.mean(p['dataset1']) -
            np.mean(d2))
    dstd = np.std(p['dataset1']) -
           np.std(d2)

    gmean = abs(dmean/
                np.mean(p['dataset1']))
    gstd = abs(dstd/
               np.std(p['dataset1']))

    rho = gmean + gstd

    return np.atleast_1d(rho)
```

and receive as input the simulated catalogue d2 and the dictionary p. Notice that the observed catalogue is contained in p. So the distance to be calculated is between p['dataset1'] and d2[10].

We must store these three functions in one file, <func_file>, and edit the sample input file provided within COSMOABC. Each keyword in the sample input file is self-explanatory, so here we only emphasise the model and prior function parameters

```
param_to_fit = mean std
param_to_sim = mean std n

mean_prior_par_name = pmin pmax
mean_prior_par_val = -2.0 4.0

std_prior_par_name = pmin pmax
std_prior_par_val = 0.1 5.0

```

[10]This format was chosen in order to optimize parallelization.

```
mean_lim = -2.0 4.0
std_lim  = 0.1 5.0

mean = 2.0
std  = 1.0
n    = 1000
...
prior_func = my_prior my_prior
```

Notice that although the variables mean and std are free parameters, we need to provide an initial numerical value, within the constraints allowed by the prior. The parameter prior_func stores the prior PDF for all the free parameters, in the sequence declared in the variable param_to_fit. Such priors do not need to follow the same family of distribution. It is possible to define a flat prior for the first parameter and a Gaussian one for the second. In that case the user input file would include, for example,

```
mean_prior_par_name = pmin  pmax
...
std_prior_par_name = pmean pstd

...

prior_func = my_prior gaussian_prior
```

considering pmean and pstd as the mean and standard deviation for the Gaussian prior of the second parameter under investigation.

3.2.1 Visualizing distance behaviour

Before we attempt to use the ABC sampler, it is important to have an idea of how our distance definition behaves for different combinations of model parameter values. COSMOABC has a tool which allows us to visually inspect the performance of our distance definition. The code randomly selects parameter values from the prior, performs the simulation and calculates the distance for each one of them. These distances are then plotted as a function of the parameter values, one parameter at a time. Ideally, the scatter of points in the $\rho \times \theta$ space should present a clear minimum in the neighbourhood of the most likely parameter value.

In order to test a personalized distance function, do

```
$ test_ABC_distance.py -i <input_file>
                       -f <func_file>
```

An example of the result of this test for the toy model we have been considering is shown in figure 3.1. Notice that the distance behaves as expected, approaching zero around the fiducial values mean=2.0 and std=1.0 and rapidly increasing as parameter values move further away.

It is worth mentioning that this procedure was implemented only to provide the user with an

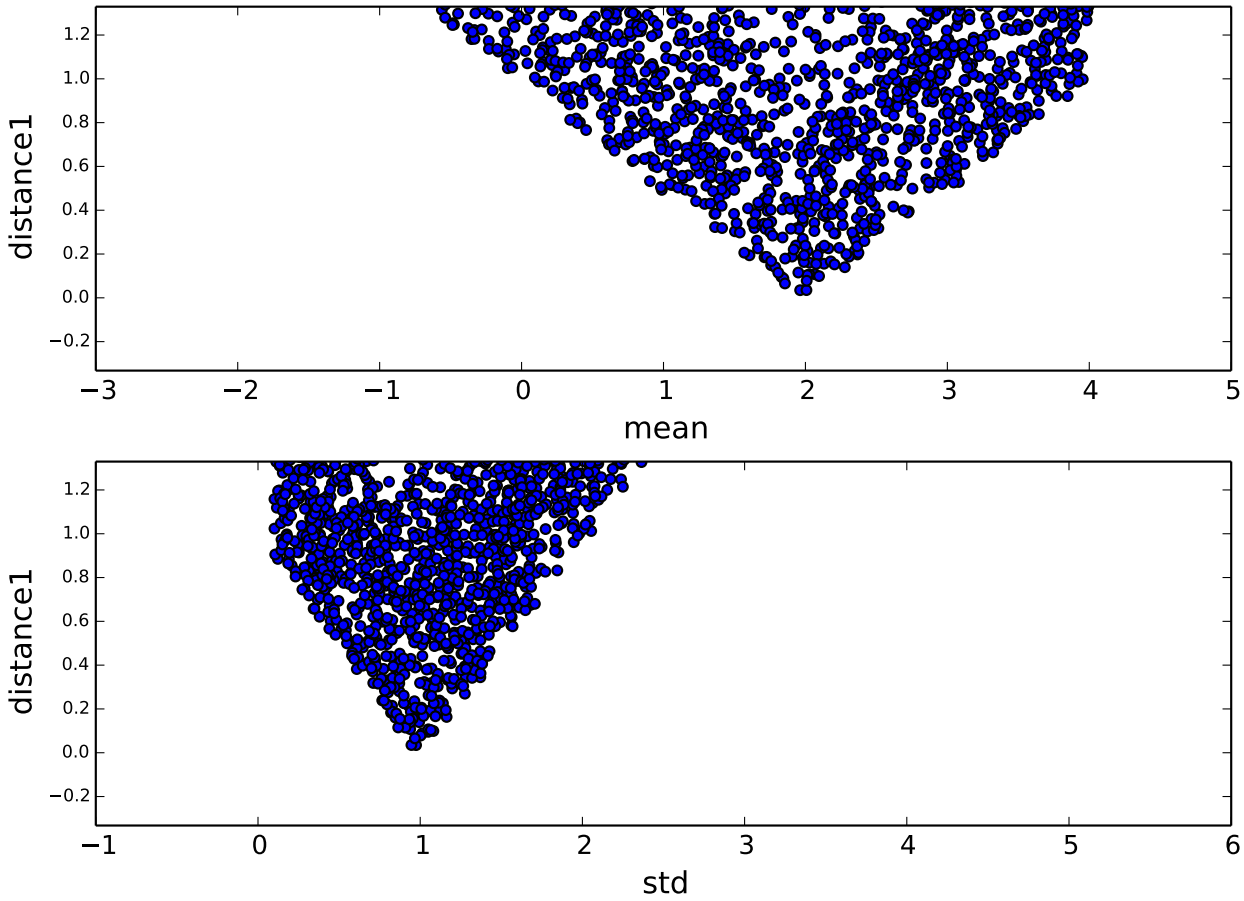


Figure 3.1: Behaviour of the distance function proposed in our toy model as a function of the free parameters mean (top) and std (bottom).

intuition regarding the distance function dependence with model parameters. The behaviour illustrated in figure 3.1 is a necessary but not sufficient characteristic of an ideal diagnostic. Selecting an appropriate distance function is an open and problem dependent challenge but it is an active area of statistical research (see e.g. [215, 217, 218]). A deeper investigation on the steps leading to an optimal distance definition, although very important, is out of the scope of this work.

3.2.2 Running the ABC sampler

After we are convinced of the performance of our distance function, we can proceed to the ABC sampler run. In COSMOABC, this is done through

```
$ run_ABC.py -i <input_file>
             -f <func_file>
```

The time necessary for the algorithm to converge depends on the efficiency of the simulator, the behaviour of the distance function and the number of particles in each particle system. We suggest an initial run with a fairly large convergence threshold, for example delta = 0.25. This means that

the code will run until it is necessary to take 4 times more draws than the number of particles in each particle system. In order to facilitate debugging and interaction with other codes, for each particle system COSMOABC outputs parameter values, distance, distance threshold, computational time and weights for each particle in ASCII tables.

Once the algorithm converges, it is possible to visualize the results with

```
$ plot_ABC.py -i <input_file>
              -f <func_file>
              -p T
```

This will generate a file containing one snapshot for each particle system from t=0 to t=T, as well as plots for the evolution of distance threshold, convergence criteria and computational time. From this first quick test, the user can either be satisfied with the achieved result or decide to continue iterating the sampler. If more iterations are required, it is only necessary to decrease the parameter delta in the user input file and continue from the last completed particle system

```
$continue_ABC.py -i <input_file>
                 -f <func_file>
                 -p T
```

3.3 Case study: cosmological parameter inference from Sunyaev-Zel'dovich surveys

The current concordance cosmology has been remarkably successful in explaining the observed properties of large scale structures [219, 220]. In this framework, the formation of such structures proceeds in a hierarchical manner driven by pressureless cold dark matter, where galaxy clusters stand out among the largest bound objects observed so far. The development of an underlying theory of cluster formation (see [221] for a review), allows us to use the abundance of clusters as well as their spatial distribution as powerful cosmological probes (e.g. [222]).

There are, however, a couple of caveats which make this an interesting problem for the ABC approach: the model is not deterministic, in the sense that it considers the observed data as a realization of a Poisson distribution (analogously to the toy model studied before) and the unavoidable modelling of the observable uncertainties and errors in both photometric redshifts and mass estimates (for a mathematical description we refer the reader to [223] and references therein). Using PMC-ABC surpasses the need to integrate a very complex likelihood function and reduces the influence of initial hypothesis on photometric redshift errors in the estimated posterior PDFs.

Since there is no previous literature on the application of PMC-ABC to this particular problem, it is crucial to establish a proof of concept. Thus, here we present results from a completely synthetic framework, where the "observed" data, $\mathcal{D}$, is one instance of our forward model. This allows us to provide a controlled scenario and to ensure our capability of recovering the input parameter values. It also facilitates the identification (and quantification) of eventual biases in the final ABC-posteriors.

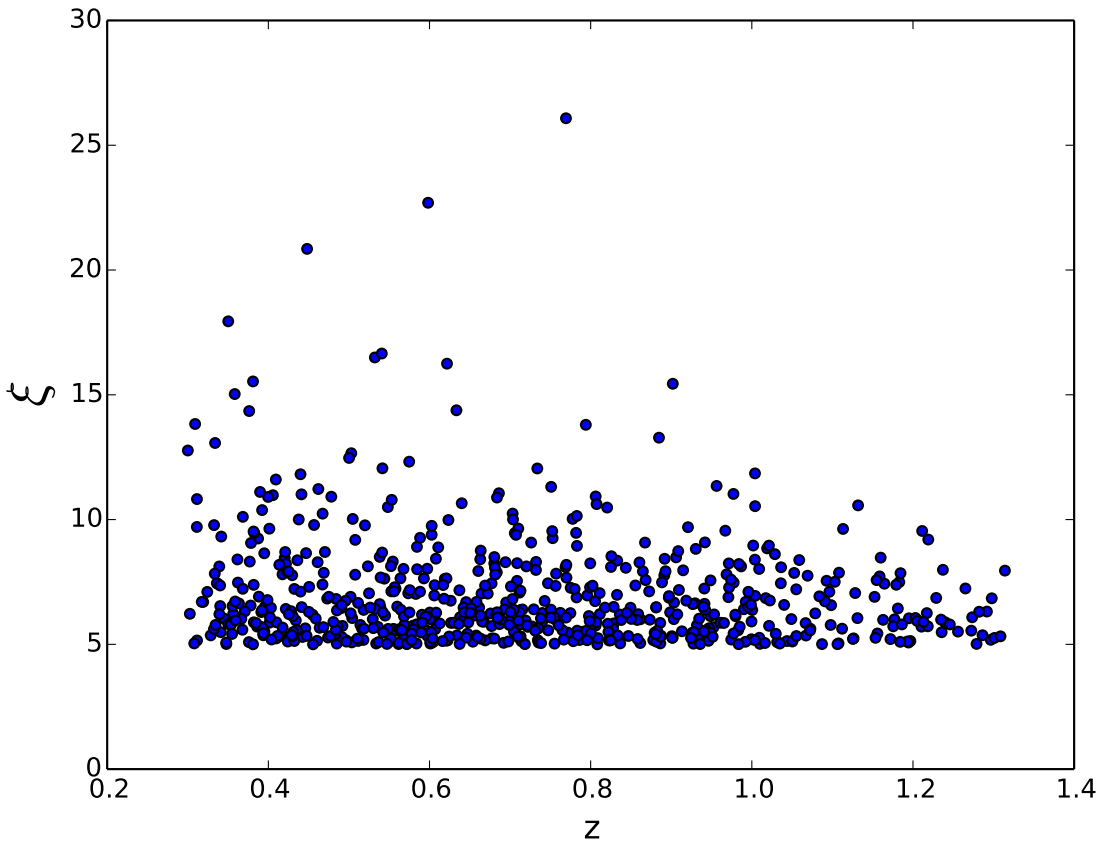


Figure 3.2: Distribution of observed features (detection significance, ξ, and redshift, z) in the "observed" catalogue, $\mathcal{D}$.

3.3.1 Simulations of the forward model

Mock catalogues were generated with the NUMCOSMO library[11], which provides a set of tools to perform cosmological calculations. The software allows a large range of possibilities for input cosmological and astrophysical parameters as well as main survey specifications (see [210] for a more detailed description). Moreover, it can also account for the presence of uncertainties from photometric redshifts and mass-observable relation (hereafter, ξ-mass relation, where ξ is the detection significance) which are crucial for a coherent analysis of galaxy cluster number counts.

The framework briefly described in section 1.4.6 allows us to relate the mean number of DM halos within a certain range of mass and redshift (eq. 1.99) to the parameters describing the underlying cosmological model. However, we still need to connect the theoretical redshift z and mass M with their equivalent observable quantities. We begin by taking into account the uncertainties from photometric redshift, z_{phot}, determination. We assume that z_{phot} follows a Gaussian distribution with mean equal to z and standard deviation $\sigma = 0.05(1+z)$, which we refer to as $P(z_{phot}|z)$.

[11] http://www.nongnu.org/numcosmo/

Thus, the expected number of clusters for a given interval of z_{phot} and M can be written as,

$$\frac{d^2N(M, z_{phot}, \boldsymbol{\theta})}{dz_{phot} d\ln M} = \int dz P(z_{phot}|z) \frac{d^2N(M, z, \boldsymbol{\theta})}{dz d\ln M}, \tag{3.4}$$

with

$$\frac{d^2N(M, z, \boldsymbol{\theta})}{dz d\ln M} dz d\ln M = f_{\text{sky}} \frac{dV(z)}{dz} \frac{dn}{dM} dz d\ln M, \tag{3.5}$$

as defined in section 1.4.6 and where $\boldsymbol{\theta}$ comprises both the cosmological and astrophysical parameters (such as those of the mass-observable relation - see for example section 2.5).

Estimating the mass enclosed in a given galaxy cluster is not a trivial task (see e.g. [224, 225] and references therein). Traditionally, one requires the recognition of indirect signatures carrying such information into observable quantities, such as optical and X-ray emissions [226, 227]. In particular, we use a mass-observable relation derived by the South Pole Telescope (SPT) team, which relies on measurements of the SZ effect (see section 2.5).

The intracluster medium (ICM) is a hot plasma which interacts with photons of the cosmic microwave background (CMB) via Compton scattering, causing a spectral distortion in the CMB radiation. This is known as the SZ effect. The integrated thermal SZ flux, Y_{SZ}, is proportional to the total thermal energy of the ICM [228, 229] and consequently it is possible to use the SZ distortions on the CMB to estimate the mass of the cluster. Due to significant uncertainties in the direct determination of Y_{SZ}, we follow here the strategy reported by the SPT [230–232], where an unbiased estimator ζ of the detection significance (signal to noise ratio) ξ is used as a mass proxy. In this context, $\zeta = \sqrt{\langle\xi\rangle^2 - 3}$ and $\zeta \propto Y_{SZ}/N_{int}$, with N_{int} denoting the noise per resolution element or the integrated noise over several resolution elements for unresolved and resolved detections, respectively. Moreover, the adopted mass scaling relation here has a different shape from the one used in section 2.5 and is given by

$$\zeta = A_{SZ} \left(\frac{M_{500}}{3 \times 10^{14} M_\odot h^{-1}} \right)^{B_{SZ}} \left(\frac{E(z)}{E(0.6)} \right)^{C_{SZ}}, \tag{3.6}$$

where $E(z) = H(z)/H_0$, $M_{500} = (4\pi/3)\, 500 \rho_{\text{crit}}\, R_{500}^3$, with $\rho_{\text{crit}} = 3H_0^2/8\pi G$ as the critical energy density, R_{500} the radius enclosing $500\times\rho_{\text{crit}}$ at the cluster redshift, and the scaling relation parameters A_{SZ} (ζ-mass normalization), B_{SZ} (ζ-mass slope) and C_{SZ} (ζ-mass redshift evolution) can be determined concomitantly with the cosmological parameters.

Finally, substituting the true mass by the unbiased estimator in eq. (3.4), the number of clusters with $\xi \in [\xi, \xi + d\xi]$ and $z_{phot} \in [z_{phot}, z_{phot} + dz_{phot}]$ can be expressed as

$$\begin{aligned} \frac{d^2N(\xi, z_{phot}, \boldsymbol{\theta})}{dz_{phot} d\xi} \quad &= \int dz P(z_{phot}|z) \\ &\int d\ln M \int d\zeta \frac{d^2N(M, z, \boldsymbol{\theta})}{dz d\ln M} \\ &P(\xi|\zeta)\, P(\ln\zeta|\ln M). \end{aligned} \tag{3.7}$$

Following [231], and [232], we assume

$$P(\ln\zeta|\ln M)d\ln\zeta = \frac{1}{\zeta\sqrt{2\pi}D_{SZ}} \times \exp\left[-\frac{(\ln\zeta - \ln M)^2}{2D_{SZ}^2}\right] d\zeta, \tag{3.8}$$

with D_{SZ} being the log-normal scatter in ζ, and

$$P(\xi|\zeta)d\xi = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\left(\xi - \sqrt{\zeta^2+3}\right)^2}{2}\right] d\xi. \tag{3.9}$$

Cosmological and astrophysical parameters for the fiducial model were chosen in accordance to [232]: $\Omega_c = 0.218$, $\sigma_8 = 0.807$, $w = -1.01$, $\Omega_{b0} = 0.044$, $H_0 = 71.15$ km/s/Mpc, $n_s = 0.97$, $A_{SZ} = 6.24$, $B_{SZ} = 1.33$, $C_{SZ} = 0.83$ and $D_{SZ} = 0.24$ (see eq. (3.6) for definitions). Telescope characteristics follow the SPT design, with minimum and maximum redshifts given by $z_{\text{min}} = 0.3$, $z_{\text{max}} = 1.32$, respectively, and survey area $\Delta\Omega = 2500$ deg^2 [233].

The simulator begins assuming that the total number of galaxy clusters with $z \in [z_{\text{min}}, z_{\text{max}}]$ and $\xi \in [\xi_{\text{min}}, \infty)$ follows a Poisson distribution. It then generates a realization of this distribution, N_{sim}, and the corresponding catalogue $\{\xi_i, z_i\}$, for $i \in \{1, N_{\text{sim}}\}$ (see [223] for more details). Here, we investigate the three-dimensional space $\{\Omega_c, \sigma_8, w\}$ with flat initial priors, $\Omega_c \in [0.01, 0.6]$, $\sigma_8 \in [0.5, 1.0]$ and $w \in [-3.0, 0.0]$. All other cosmological parameters are considered known and fixed at the values reported above.

COSMOABC contains a wrap of the NUMCOSMO simulator which can be accessed through the user input file keyword

```
simulation_func = numcosmo_sim_cluster
```

and an example of the input file with all other options tailored for NUMCOSMO simulations is also provided within the package.

Figure 3.2 displays the static simulated catalogue we used as “observed” data in the $\xi \times z$ sample space. The sample is composed by 671 clusters with $z \in [0.30, 1.32]$ and $\xi \in [5, 26]$.

3.3.2 Distance

The complexity enclosed in the cosmological simulations requires some sophistication in designing the distance function. COSMOABC has two built-in definitions which proved to be effective in the galaxy cluster counts scenario: quantiles and Gaussian radial basis function (GRBF) distances.

Let us start by briefly describe a distance defined in terms of Gaussian radial basis functions (GRBF), ρ_{GRBF} (within COSMOABC the function is called distance_GRBF). This is not a distance in the mathematical sense, since $\rho_{\text{GRBF}}(\mathcal{D}, \mathcal{D}_S) \neq \rho_{\text{GRBF}}(\mathcal{D}_S, \mathcal{D})$, but as the PMC-ABC is centred

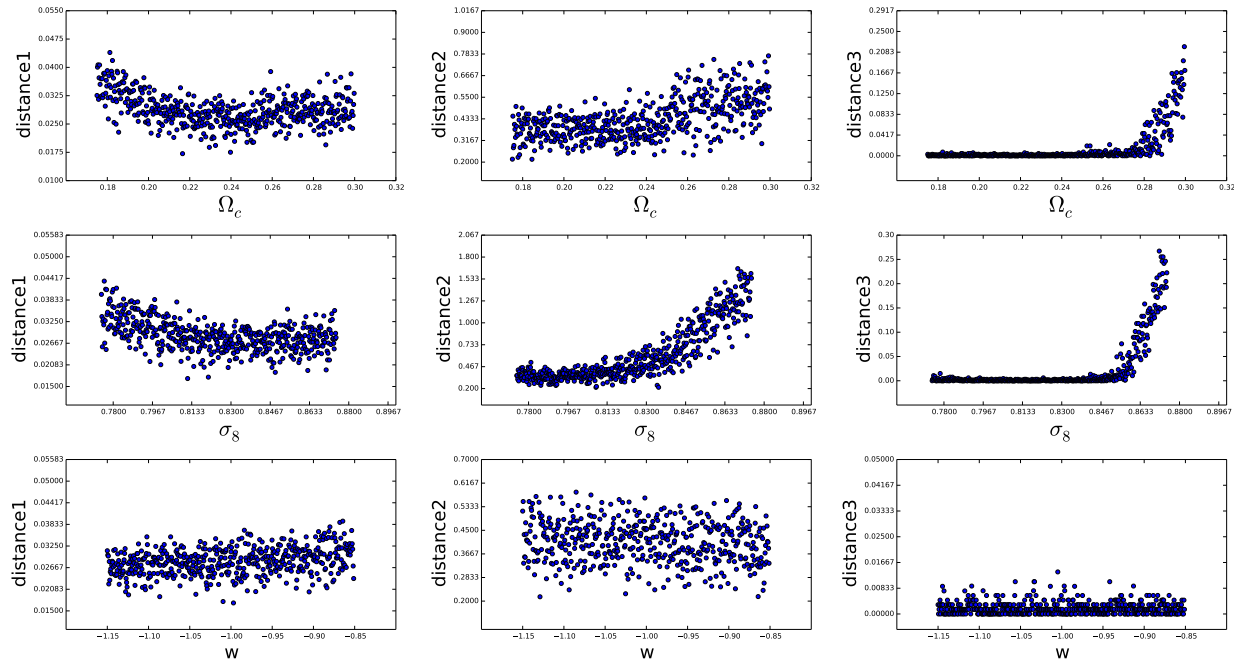


Figure 3.3: Behaviour of the quantile distance function in the context of galaxy cluster number counts. Each panel illustrates how the elements of the quantile based distance vary as a function of the cosmological parameters for 10^4 random draws from the prior. Lines run through parameters and columns through distance elements.

in the observed catalogue, it is enough to guide the sampler to the correct posterior distribution over the evolution of particle systems.

For a given simulated sample, the approximation of its underlying model is computed using a GRBF interpolation [234],

$$G_{\mathsf{R}}(\xi, z|\mathcal{D}) = \sum_{i=1}^{N_{\mathsf{sim}}} \frac{1}{2\pi\sqrt{\mathsf{det}(\boldsymbol{C})}} \exp\left(-\frac{\boldsymbol{d}_i^T \cdot \boldsymbol{C}^{-1} \cdot \boldsymbol{d}_i}{2}\right), \tag{3.10}$$

where

$$\begin{aligned} \boldsymbol{d}_i &= (\xi - \xi^i_{\mathcal{D}_S}, z - z^i_{\mathcal{D}_S}), \\ \boldsymbol{C} &= s^2 \mathsf{cov}(\mathcal{D}), \end{aligned}$$

s is a scale parameter and $\mathsf{cov}(\mathcal{D})$ is the covariance matrix of the observed data. Each element of the sum in eq. (3.10) works as a kernel density distribution centred in the $i-th$ simulated cluster with $(\xi_i,\ z_i)$. Consequently, $G_{\mathsf{R}}(\xi_*, z_*)$ is the number of clusters we expect to observe having (ξ_*, z_*).

We can re-obtain the total number of objects in the catalogue through

$$\int \mathsf{d}\xi \int \mathsf{d}z G_{\mathsf{R}}(\xi, z|\mathcal{D}) = N_{\mathsf{sim}}. \tag{3.11}$$

Making use of the auxiliary function

$$\ln f(\mathcal{D}|\mathcal{D}_S) = \sum_{j=1}^{N} \ln(G_{\mathsf{R}}(z_j, \xi_j|\mathcal{D}_S)) - N_{\mathsf{sim}}, \tag{3.12}$$

with the index j running through all the data points in $\mathcal{D}$ [223], the distance between the "observed" and simulated catalogues is given by

$$\rho(\mathcal{D}, \mathcal{D}_S) = -2 \ln \left(\frac{f(\mathcal{D}|\mathcal{D}_S)}{f(\mathcal{D}|\mathcal{D})} \right). \tag{3.13}$$

Therefore, in each iteration, t, we will only accept those parameter values whose forward model satisfy $\rho_t(\mathcal{D}, \mathcal{D}_S) \leq \epsilon_t$ and the ABC-PMC algorithm can be employed normally[12].

The scale parameter s (eq. (3.11)) regulates our tolerance towards distinct distributions which produce the same total number of objects in a catalogue. Suppose we begin with a large ϵ_1 and follow the ABC-PMC algorithm reducing ϵ_t at each time-step. If s is too small the probability of finding parameter values satisfying the distance threshold will drop steadily, rendering the algorithm unable to further reduce ϵ_t. On the other hand, if s is too large, the density function will evolve to a very flat behaviour losing all information about the underlying distribution of $\mathcal{D}$. Thus, s must be chosen such that most of the shape information in $\mathcal{D}$ is retained, while still being feasible to reduce ϵ_t until the desired precision is achieved. For the specific case outlined here, we found that any $s \in [0.1, 0.5]$ will lead to well constrained and unbiased results.

In order to call this distance definition from within COSMOABC, the input file must include the extra parameter s and the function definition

```
...
distance_func = distance_GRBF
s   = 0.15
...
```

The GRBF distance is more time consuming, since it takes into account the correlation between the observed features. However, it might be worth to use it in highly correlated data scenarios. The current version of the NUMCOSMO library includes an ABC sampler using the GRBF distance.

The distance_quantile function, in turn, returns a vector $\boldsymbol{\varrho}$, having $L+1$ dimensions, where L is the number of measured features[13]. For each feature (column in the catalogue), it calculates a few equally spaced quantiles[14]. At every quantile, the values of the cumulative distribution functions (CDF) coming from simulated and observed catalogues are subtracted and the square root of their sum is returned. The last dimension accounts for the variability in the total number of objects. If l

[12] Note that the denominator in eq. (3.13) is a scalar independent of the mock data. This ensures the distance function convergence.

[13] In our case, $L = 2$, for observed features ξ and redshift.

[14] The number of quantiles if defined by the user in the input file, through the keyword quantile_nodes.

is the number of objects in $\mathcal{D}$ and l_S is the number of objects in $\mathcal{D}_S$, the last element of $\boldsymbol{\rho}$ will be

$$\rho_{-1} = \text{max}\left[\text{abs}\left(1 - \frac{l}{l_S}\right), \text{abs}\left(1 - \frac{l_S}{l}\right)\right]. \tag{3.14}$$

In the construction of the first particle system, the magnitude of this vector, $|\boldsymbol{\rho}|$, is used to select the set of N particles with smaller distances. Once the first particle system is constructed, the distance threshold $\boldsymbol{\epsilon}$ will also be a $L+1$-dimensional vector. A new set of parameter values $\boldsymbol{\theta}$ will only be accepted to populate the next particle system if it satisfies the 3 distance thresholds independently.

We emphasise that the this is only a simple and computationally fast distance definition which proved to be efficient in this synthetic scenario of cosmological inference from galaxy cluster number counts for the illustrative purposes of this work. Whenever using ABC in a real data situation, the user must design a distance function which preserves these features for the problem at hand (e.g., see section 3.3 of [199]).

Figure 3.3 illustrates the effectiveness of this distance definition in determining the cosmological parameters based on SZ flux measurements. The distance calculations were performed using the COSMOABC tool described in section 3.2.1, however, in order to make the visualization lighter, we display binned results in all three free parameters. In each panel the horizontal axis was divided in 500 bins and each dot represents the smallest distance found in that bin for 10^4 draws. From figure 3.3 we see that the first (comparison of CDF over redshift) and second (comparison of CDF over ξ) distance elements do present a local minimum around the fiducial values for Ω_m and σ_8, although the behaviour is much lighter than in the previously discussed toy model (figure 3.1). The role of the third element (comparison between the total number of objects) is to impose an upper limit on the free parameter values, since this element increases steadily for $\Omega_{m0} \geq 0.28$ and $\sigma_8 \geq 0.86$. We also see that there is little hope in using this distance to constraint w, since there is no significant change in behaviour for the three distance elements.

3.3.3 Results

Specific tools are also available for the case of a SZ survey using COSMOABC. Once all the choices are made in the user input file it is possible to run the ABC sampler using

```
$ run_ABC_NumCosmo.py -i <input_file>
```

Analogously, if the user is interested in continuing the calculations from a given particle system T, this can be done using

```
$ continue_ABC_NumCosmo.py
                -i <input_file>
                -p T
```

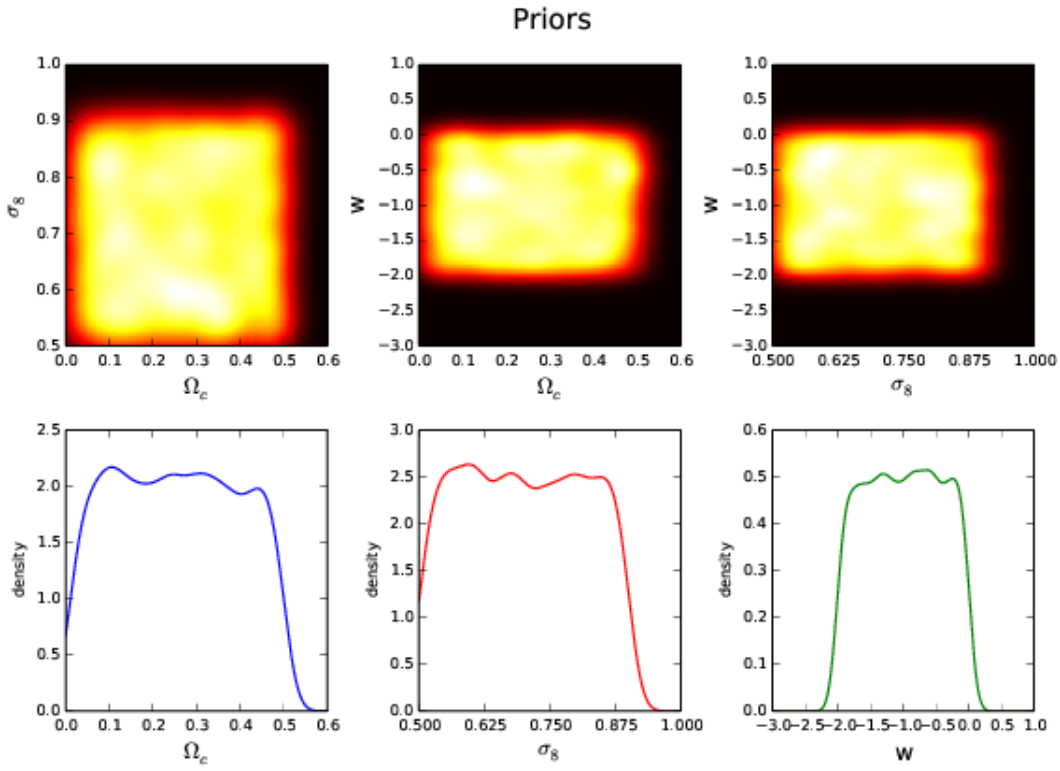
In case a user defined distance or prior is chosen, it is necessary to include the -f option followed by

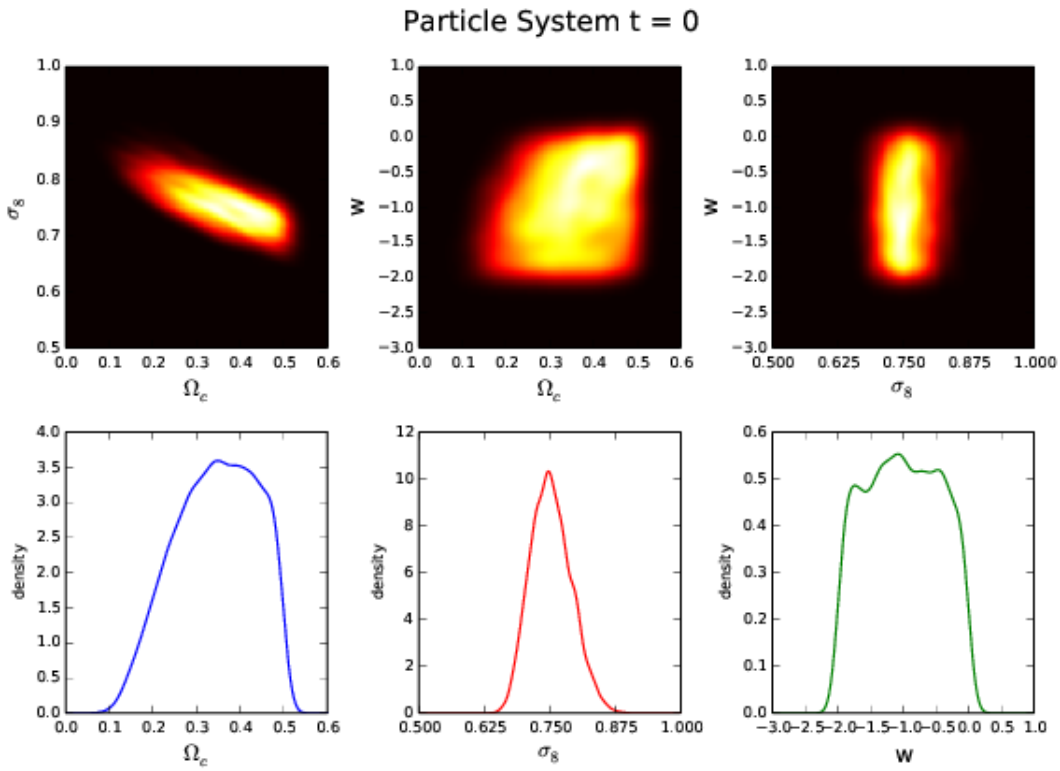
the name of the function file in both examples above. Plots can be generated as shown in section 3.2.2.

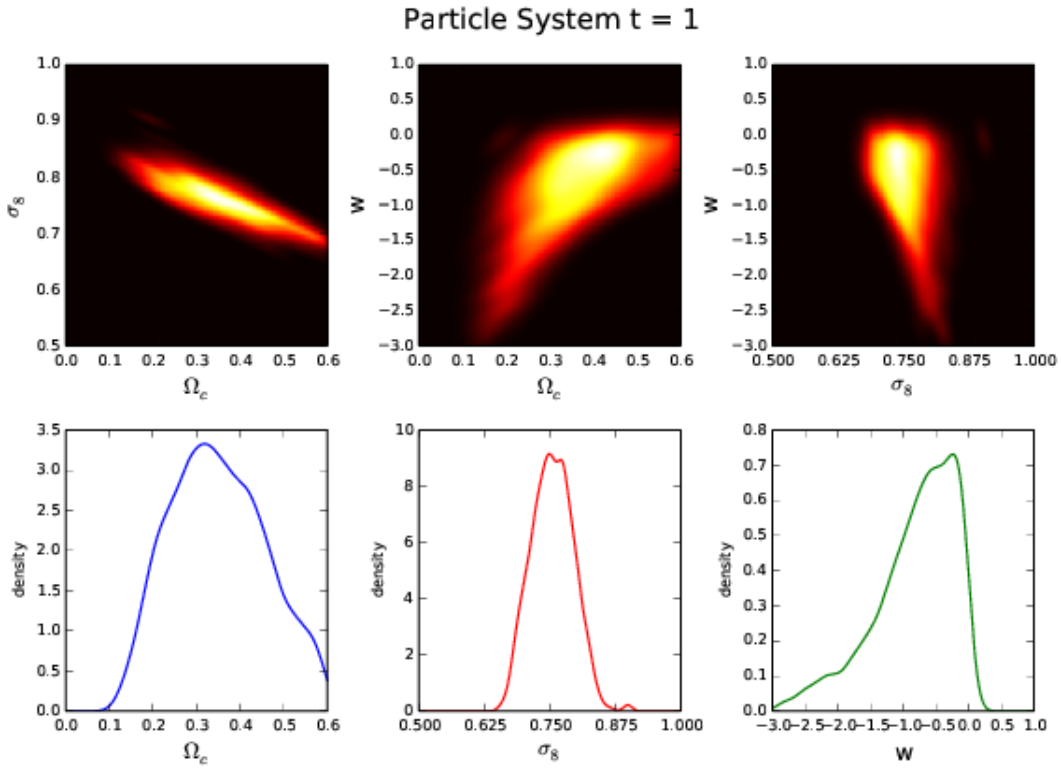
Credible intervals from the ABC-PMC estimated posterior distributions, using the distance_quantile function alone, are shown in figure 3.1. The upper panels show 2-dimensions posteriors over the free parameters and the bottom panels display the profile of each parameter individually. Frames show the evolution of the approximated posteriors for consecutive particle systems. The first frame merely represents the initial prior: a flat PDF over all the free parameters. The next frame displays results from the first particle system ($t = 0$), where we generate a large number of simulations ($M = 50000$) and kept only the 10% with smaller distance. From $t = 1$ we clearly see how the posterior evolves and adapts through subsequent iterations. The credible intervals not only shrink, but also become asymptotically well behaved for further particle systems.

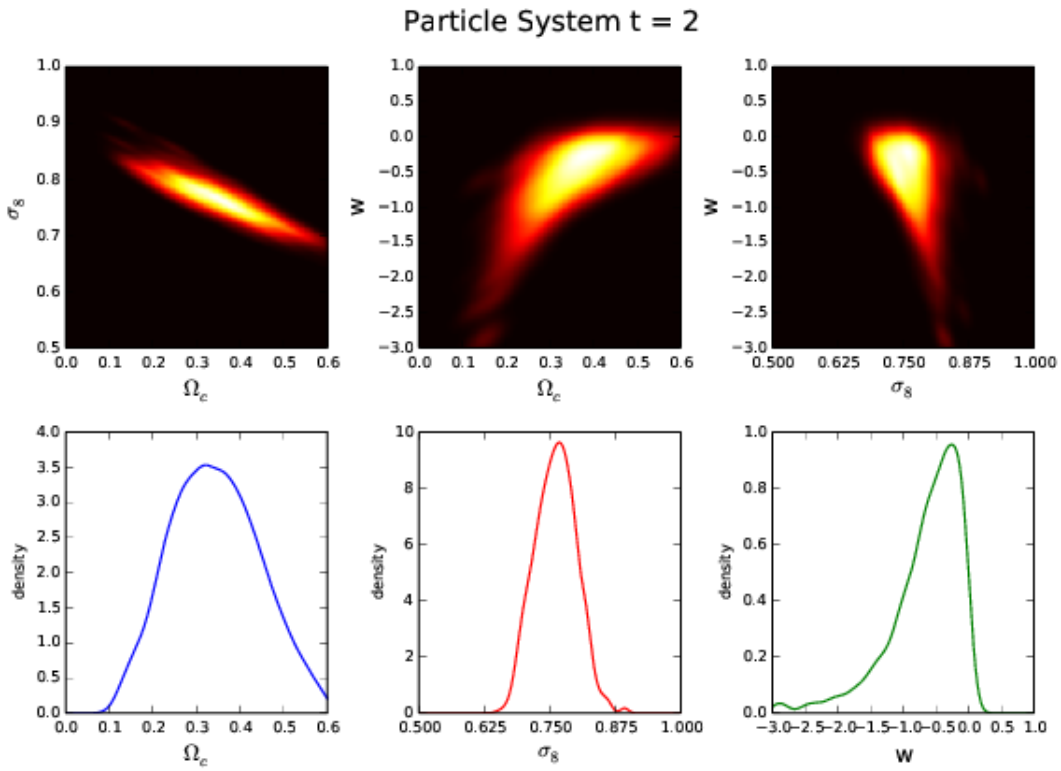
Worth noting, for this example each particle system holds $N = 5000$ particles and the convergence was achieved in 9 iterations for delta = 0.01. In case a tighter posterior is desirable for Ω_c and σ_8, one can simply decrease the convergence criteria, letting the system evolve for a little longer. If further information on w is desired, a more informative distance definition should also be used[15] (see figure 3.3). The evolution of the distance threshold and convergence criteria are shown in figures 3.2 and 3.3, respectively.

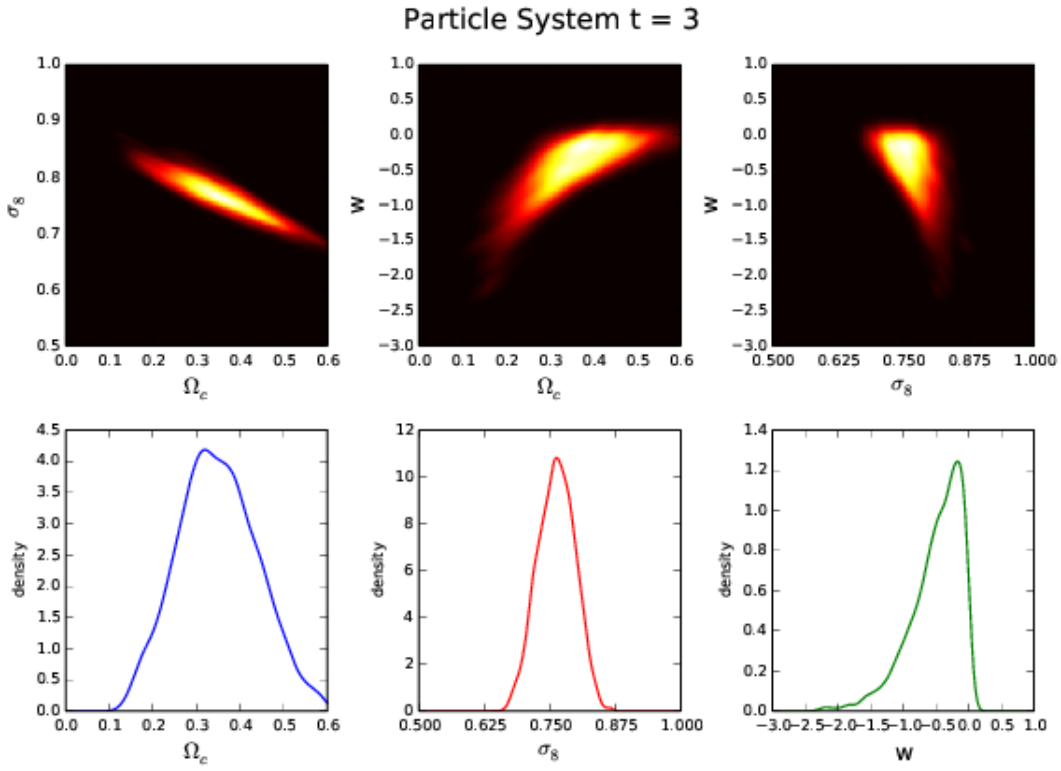
[15]The quantile distance was chosen due to its simplicity and low computational cost. COSMOABC also contains a distance definition in eq. (3.10) which accounts for potential correlations between two parameters in a catalogue. We advise the user to consider the GRBF distance as well as the combination with other cosmological probes in case tight intervals over w are desired.

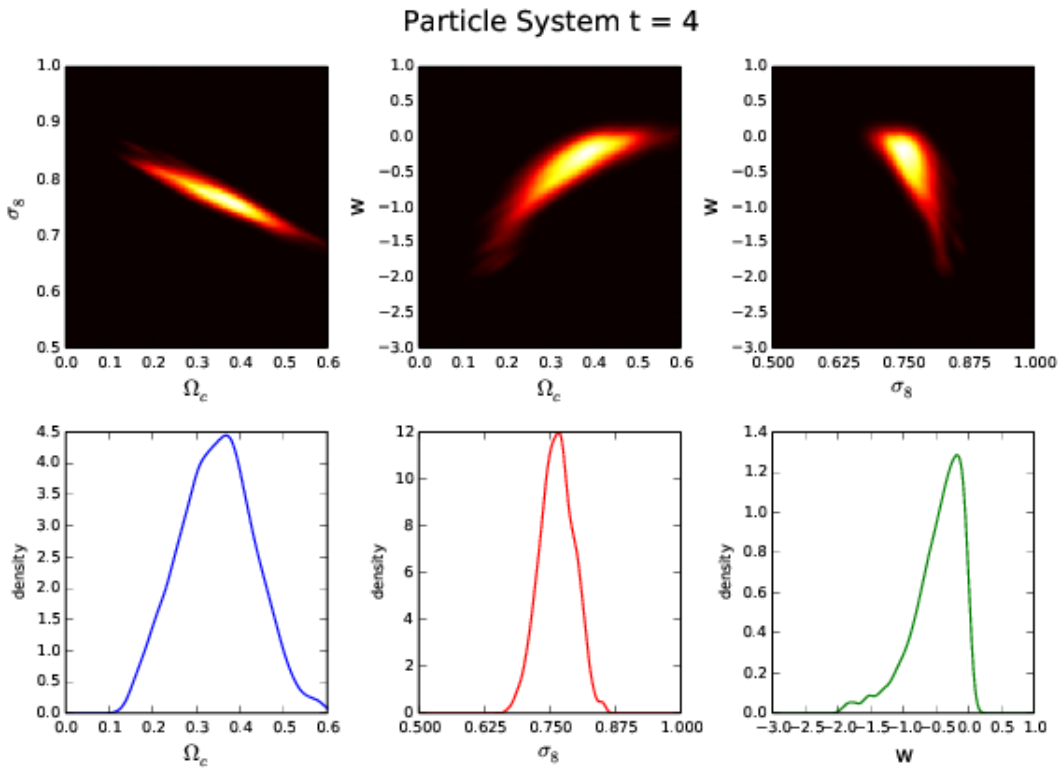

Priors
σ_8
Ω_c
w
density

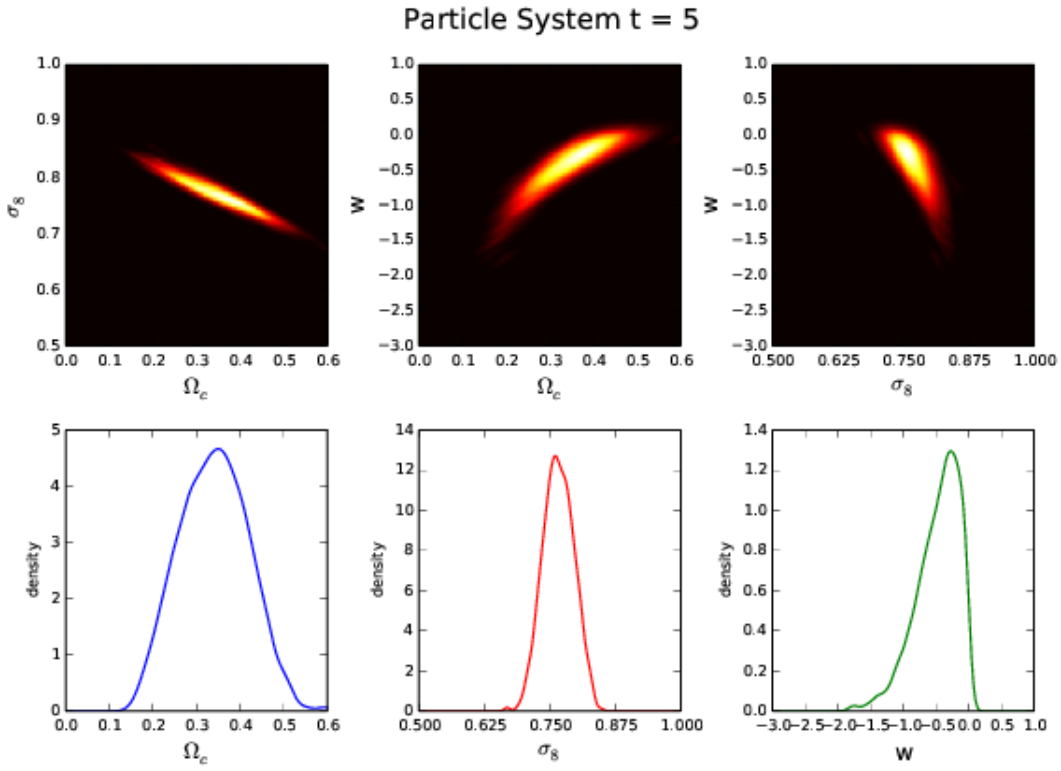

Particle System t = 0
σ_8
Ω_c
w
density

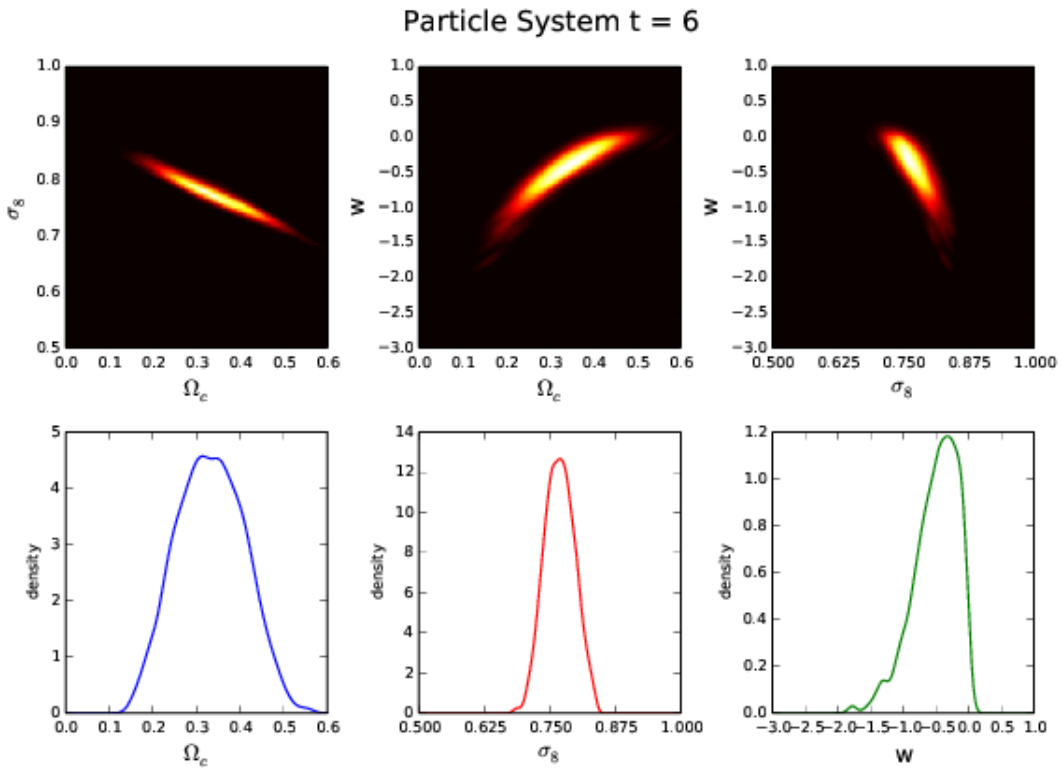

Particle System t = 1
σ_8
w
Ω_c
density


Particle System t = 2
σ_8
w
Ω_c
density


Particle System t = 3
σ_8
w
Ω_c
σ_8
density
w


Particle System t = 4
σ_8
w
Ω_c
σ_8
density
w


Particle System t = 5
σ_8
w
w
Ω_c
Ω_c
σ_8
density
density
density
Ω_c
σ_8
w


Particle System t = 6
σ_8
w
w
Ω_c
Ω_c
σ_8
density
density
density
Ω_c
σ_8
w

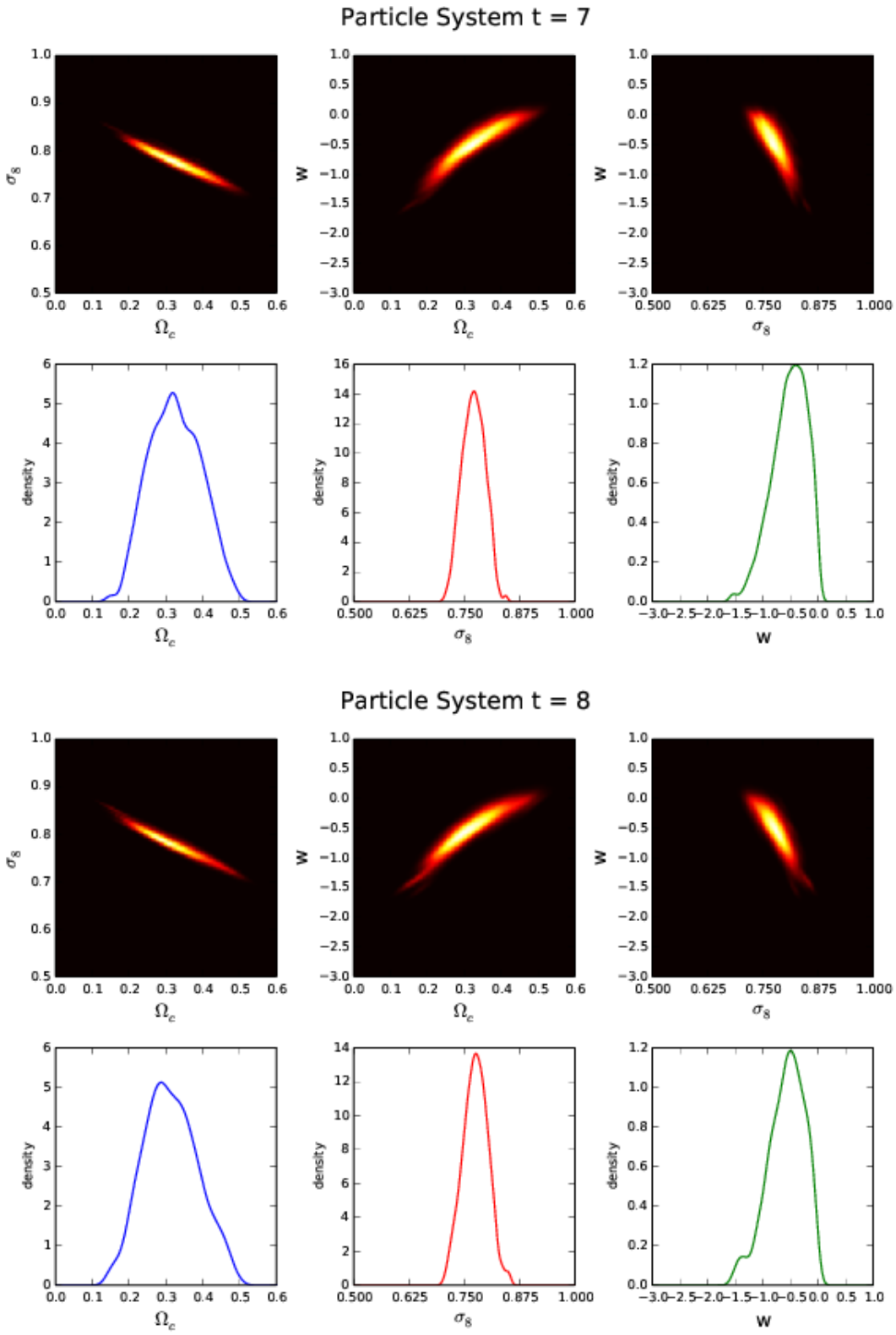


Figure 3.1: Results from coupling COSMOABC to the NUMCOSMO simulator. Frames run from successive iterations of the PMC-ABC algorithm. **Upper panel**: two-dimensional representation of the ABC posteriors in each iteration. **Lower left panel**: evolution of the dark matter density profile. **Lower centre panel**: evolution of the posterior over σ_8. **Lower right panel**: evolution of the PDF profile over the dark energy equation of state parameter.

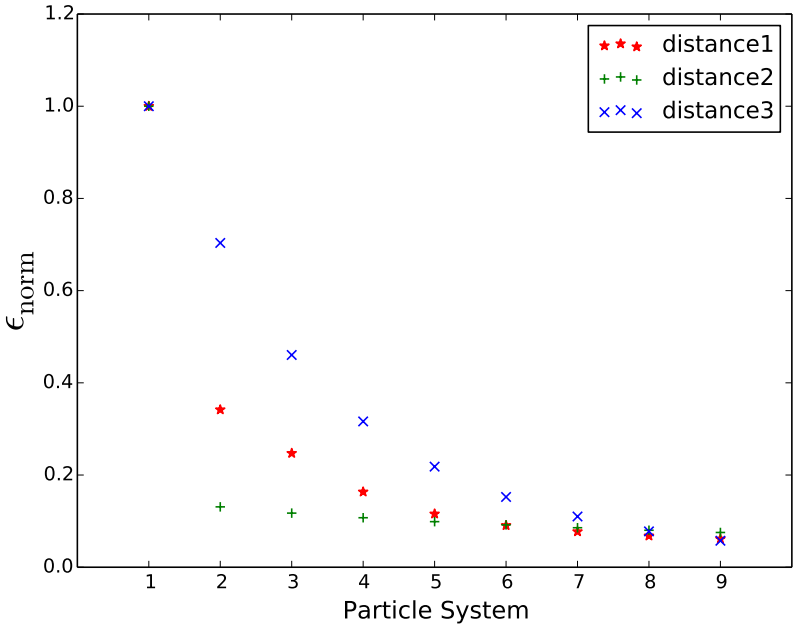


Figure 3.2: Evolution of the distance threshold. The first (stars), second (+) and third (x) elements of the quantile distance function were normalized by their respective larger values. The horizontal axis runs through all the particle systems shown in figure 3.1.

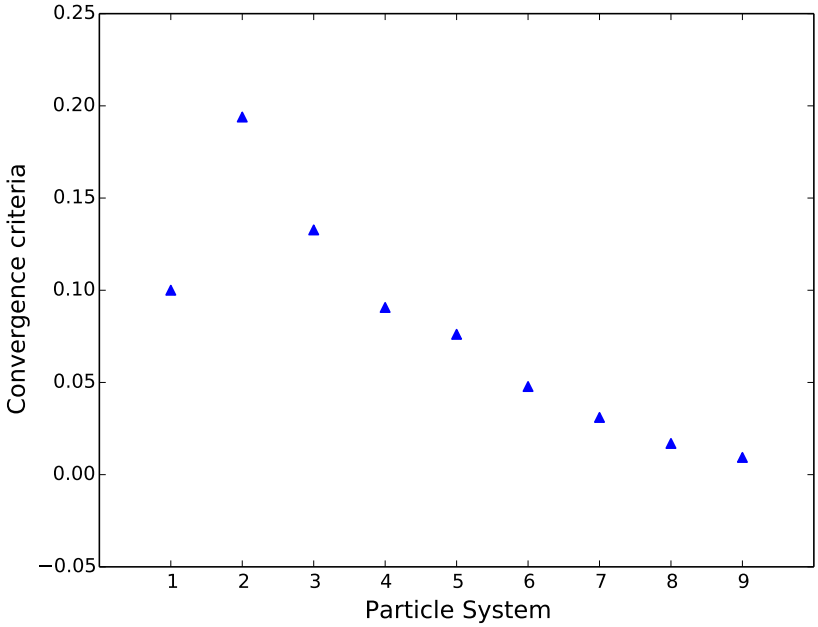


Figure 3.3: Evolution of the convergence criteria for the results shown in figure 3.1.

3.4 Conclusions

In this chapter, we have presented COSMOABC, a Python implementation of Population Monte Carlo Approximate Bayesian Computation (PMC-ABC) algorithm with adaptive importance sampling. Traditional methods of parameter inference are useful if the likelihood is available and feasible to compute. Due to the increasing amount of data and their complex modelling in all areas of astronomy and cosmology, more and more computational power is required in order to explore larger parameter spaces whose internal correlations can often be impractically complicated or unknown. Thus, obtaining a statistical tool which bypasses the need of fully evaluating the likelihood is imperative. PMC-ABC presents an interesting alternative. COSMOABC is the first such implementation targeted to the astronomy and cosmology community.

The cosmological simulations are done through a connection with the *Numerical Cosmology* library (NUMCOSMO), but the code is flexible enough for user-specified distance, simulation and prior functions.

We stress that ABC is not a substitute for standard MCMC algorithms when the likelihood is completely known or easy to calculate. It is a viable alternative when we are not able to handle the likelihood itself, and thus in situations where a MCMC is not feasible.

We have demonstrated the power of COSMOABC in estimating posterior probability distributions in two situations: a simple toy model and a complex cosmological simulation of Sunyaev-Zel'dovich survey. In both cases, we demonstrated how COSMOABC allows a good approximation of the true posterior probability distribution with a fairly simple and user-friendly interface. We used a completely synthetic environment in order to demonstrate the efficiency of the method and to be able to address the accuracy of the results. We hope this will be useful not only to cosmologists, but to all research areas in astronomy where simulations are becoming increasingly more accessible and systematics are making likelihood functions even more intractable.

Chapter 4

Summary and Final Remarks

Primordial non-Gaussianities have become a very active area of research within Modern Cosmology, since they constitute a powerful observational probe that may help us discriminate between the different inflationary scenarios. Although the simplest inflation models, i.e. single field slow-roll inflation, predict that primordial fluctuations follow a nearly Gaussian statistics, more complex models may produce larger deviations from Gaussian initial conditions. In this thesis we have investigated how the estimation of cosmological parameters using galaxy clusters counts, may be affected by wrongly assuming that no primordial non-Gaussianities exist. First, we focused our attention on the dark energy equation of state parameter only. We found that ignoring primordial non-Gaussianities in our analysis induces an apparent evolution with time of the effective dark energy equation of state, which is translated in an evident discontinuity. This discontinuity constitutes a new diagnostic for the existence of primordial non-Gaussianities.

Encouraged by the previous results, we also estimated the magnitude of the biases that may be introduced in the estimation of a broader set of cosmological parameters by assuming the absence of primordial non-Gaussianities. The parameters considered were the present-day dark energy energy density Ω_{de0}, the dark energy equation of state parameter w and the amplitude of the primordial fluctuations σ_8 and, again, we resorted to galaxy clusters number counts as probes. The results show that the biases induced in the cosmological parameters due to wrongly assume Gaussian initial conditions, although small for Ω_{de0} and w, are quite significant for σ_8. If the non-linear parameter f_{NL} that controls the level of non-Gaussianity is scale-independent, then in light of the constraints obtained by the Planck team, we can safely neglect primordial non-Gaussianities in the determination of cosmological parameters using the abundance of galaxy cluster. This is no longer necessarily true, if f_{NL} varies with scale. Furthermore, if a significant level of non-Gaussianity is present at clusters scales, order of $f_{NL} \approx -240$, then our results indicate that it would be possible to close the gap between the two conflicting constraints on σ_8, obtained by the Planck team.

Even though galaxy clusters constitute a reservoir of valuable cosmological information, retrieving it is only possible if an accurate and reliable estimation of the masses of these objects can be done. From the observational standpoint, this is a very difficult task and often one has to rely on intrinsic statistical properties of clusters as mass proxies. Thus it becomes important to assess

up to what point primordial non-Gaussianities may influence the determination and calibrations of the galaxy clusters scaling relations. We performed a series of N-body simulations with different levels of non-Gaussian initial conditions and, for the first time, investigated the impact of primordial non-Gaussianities on the scaling of the mass weighted temperature T_{mw}, entropy S, core X-ray bolometric luminosity Y_X, and Compton parameter Y with the mass, M. We have found that in general, the scaling's power-law index α shows no systematic variation with f_{NL}. On the other hand, the effects are much more severe on the normalization $\log(A)$ and the redshift evolution β of the scaling relations.

Modern Cosmology is now a data driven field and thus is becoming less limited by the amount of high precision observational data available. Therefore, it is fundamental to develop tools and sophisticated algorithms to retrieve all the relevant cosmological information from the available observations. In this regard, the final part of this thesis introduces the COSMOABC package, which is a Python implementation of the Population Monte Carlo Approximated Bayesian Computation algorithm with adaptive importance sampling. The package has the virtue of being highly parallelized and very flexible, so it can be coupled with external codes. COSMOABC was applied to a problem of interest in cosmology: the estimation of the posterior probability distribution of cosmological parameters using a complex simulation of a galaxy cluster Sunyaev–Zeldovich survey. COSMOABC yielded a very good approximation of the true posterior probability distribution, while achieving convergence in a very short period of time.

The scientific studies conducted in this thesis on primordial non-Gaussianities were done using only the abundance of galaxy clusters as the source of cosmological information. However, the scope of observables that can be used to probe Primordial non-Gaussianities is much broader. Although current constraints on the level of non-Gaussianity from CMB data [34, 35] are consistent with Gaussian initial conditions, it is still very relevant to continue with the search of a non-Gaussian signature in the primordial density field. One plausible justification relies on the fact that the non-linear parameter f_{NL} could be scale dependent and thus primordial non-Gaussianities may exhibit a higher positive/negative amplitude on smaller scales. This makes the clustering of galaxies and galaxies clusters the observable of choice to probe for non-Gaussianties (see e.g. [235–238]). However, at smaller scales non-Gaussianities due to the non-linear gravitational evolution of structures could eventually mask the primordial non-Gaussian features of the density field. The EUCLID-ESA mission that is currently being planned, will measure accurately the positions of tens of millions of galaxies, which will be used for galaxy clustering. The EUCLID mission will also provide data on the shape distortions of galaxies induced by gravitational lensing effects of dark matter, whose statistical properties have also been shown to be sensitive to primordial non-Gaussianities (see e.g. [39, 40, 239–242]). Primordial non-Gaussianities may also have a detectable impact in the 21 cm radiation due to spin-flip transitions of neutral hydrogen during the epoch of reionization or earlier during the cosmic Dark Ages (see e.g. [41, 243–245]), which could provide the same level or better constraints on f_{NL} than the Planck collaboration for a SKA like survey [244]. Whatever probe or combination of probes we may use to search for primordial non-Gaussianities, the fact is that they offer a unique and valuable way to learn about inflation, which in turn teaches us about the interaction and inner workings of quantum fields in the early Universe.

Bibliography

[1] A. Einstein. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*

[2] A. Einstein. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152.*, 1917.

[3] W. de Sitter. On the relativity of inertia. Remarks concerning Einstein's latest hypothesis. *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences*, 19:1217–1225, 1917.

[4] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, 1922.

[5] A. Friedmann. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21(1):326–332, 1924.

[6] G. Lemaître. Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. MNRAS, 91:483–490, March 1931.

[7] H. P. Robertson. On the Foundations of Relativistic Cosmology. *Proceedings of the National Academy of Science*, 15:822–829, November 1929.

[8] H. P. Robertson. Kinematics and World-Structure. ApJ, 82:284, November 1935.

[9] A. G. Walker. On the formal comparison of Milne's kinematical system with the systems of general relativity. MNRAS, 95:263–269, January 1935.

[10] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929.

[11] R. A. Alpher and R. Herman. Evolution of the Universe. Nature, 162:774–775, November 1948.

[12] G. Gamow. The Origin of Elements and the Separation of Galaxies. *Physical Review*, 74:505–506, August 1948.

[13] G. Gamow. The Evolution of the Universe. Nature, 162:680–682, October 1948.

[14] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XVI. Cosmological parameters. A&A, 571:A16, November 2014.

[15] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. ApJ, 142:419–421, July 1965.

[16] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic Black-Body Radiation. ApJ, 142:414–419, July 1965.

[17] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. ApJ, 147:73, January 1967.

[18] P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. ApJ, 162:815, December 1970.

[19] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss, and D. T. Wilkinson. Structure in the COBE differential microwave radiometer first-year maps. ApJ, 396:L1–L5, September 1992.

[20] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. AJ, 116:1009–1038, September 1998.

[21] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and T. S. C. Project. Measurements of Ω and Λ from 42 High-Redshift Supernovae. ApJ, 517:565–586, June 1999.

[22] http://www.2dfgrs.net/.

[23] http://www.sdss.org/.

[24] http://www.roe.ac.uk.

[25] https://www.sdss3.org.

[26] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness problems. Phys. Rev. D, 23:347–356, January 1981.

[27] P. Creminelli. On non-Gaussianities in single-field inflation. J. Cosmology Astropart. Phys., 10:3, October 2003.

[28] J. Maldacena. Non-gaussian features of primordial fluctuations in single field inflationary models. *Journal of High Energy Physics*, 5:13, May 2003.

[29] D. H. Lyth and Y. Rodríguez. Inflationary Prediction for Primordial Non-Gaussianity. *Physical Review Letters*, 95(12):121302, September 2005.

[30] D. Seery and J. E. Lidsey. Primordial non-Gaussianities in single-field inflation. J. Cosmology Astropart. Phys., 6:3, June 2005.

[31] E. Sefusatti and E. Komatsu. Bispectrum of galaxies from high-redshift galaxy surveys: Primordial non-Gaussianity and nonlinear galaxy bias. Phys. Rev. D, 76(8):083004, October 2007.

[32] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18, February 2011.

[33] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208:19, October 2013.

[34] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XXIV. Constraints on primordial non-Gaussianity. A&A, 571:A24, November 2014.

[35] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, and et al. Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. *ArXiv e-prints*, February 2015.

[36] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan. Constraints on local primordial non-Gaussianity from large scale structure. J. Cosmology Astropart. Phys., 8:31, August 2008.

[37] S. Matarrese and L. Verde. The Effect of Primordial Non-Gaussianity on Halo Bias. ApJ, 677:L77–L80, April 2008.

[38] T. Giannantonio, C. Porciani, J. Carron, A. Amara, and A. Pillepich. Constraining primordial non-Gaussianity with future galaxy surveys. MNRAS, 422:2854–2877, June 2012.

[39] B. M. Schäfer, A. Grassi, M. Gerstenlauer, and C. T. Byrnes. A weak lensing view on primordial non-Gaussianities. MNRAS, 421:797–807, March 2012.

[40] S. Hilbert, L. Marian, R. E. Smith, and V. Desjacques. Measuring primordial non-Gaussianity with weak lensing surveys. MNRAS, 426:2870–2888, November 2012.

[41] H. Tashiro and S. Ho. Constraining primordial non-Gaussianity with CMB-21 cm cross-correlations? MNRAS, 431:2017–2023, May 2013.

[42] Y. Takeuchi, K. Ichiki, and T. Matsubara. Application of cross correlations between CMB and large-scale structure to constraints on the primordial non-Gaussianity. Phys. Rev. D, 85(4):043518, February 2012.

[43] S. Matarrese, L. Verde, and R. Jimenez. The Abundance of High-Redshift Objects as a Probe of Non-Gaussian Initial Conditions. ApJ, 541:10–24, September 2000.

[44] J. Robinson and J. E. Baker. Evolution of the cluster abundance in non-Gaussian models. MNRAS, 311:781–792, February 2000.

[45] E. Sefusatti, C. Vale, K. Kadota, and J. Frieman. Primordial Non-Gaussianity and Dark Energy Constraints from Cluster Surveys. ApJ, 658:669–679, April 2007.

[46] M. LoVerde, A. Miller, S. Shandera, and L. Verde. Effects of scale-dependent non-Gaussianity on cosmological structures. J. Cosmology Astropart. Phys., 4:14, April 2008.

[47] N. Dalal, O. Doré, D. Huterer, and A. Shirokov. Imprints of primordial non-Gaussianities on large-scale structure: Scale-dependent bias and abundance of virialized objects. Phys. Rev. D, 77(12):123514, June 2008.

[48] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga. Probing primordial non-Gaussianity with large-scale structure. Phys. Rev. D, 69(10):103513, May 2004.

[49] H. Mathis, J. M. Diego, and J. Silk. The case for non-Gaussianity on cluster scales. MNRAS, 353:681–688, September 2004.

[50] M. Kamionkowski, L. Verde, and R. Jimenez. The void abundance with non-gaussian primordial perturbations. J. Cosmology Astropart. Phys., 1:010, January 2009.

[51] T. Y. Lam, R. K. Sheth, and V. Desjacques. The initial shear field in models with primordial local non-Gaussianity and implications for halo and void abundances. MNRAS, 399:1482–1494, November 2009.

[52] G. D'Amico, M. Musso, J. Noreña, and A. Paranjape. Excursion sets and non-Gaussian void statistics. Phys. Rev. D, 83(2):023521, January 2011.

[53] T. Sekiguchi and S. Yokoyama. Void bias from primordial non-Gaussianities. *ArXiv e-prints*, April 2012.

[54] B.F. Schutz. *A First Course in General Relativity*. Series in physics. Cambridge University Press, 1985.

[55] A. R. Liddle and D. H. Lyth. *Cosmological Inflation and Large-Scale Structure*. June 2000.

[56] P.P. Coles and F. Lucchin. *Cosmology: The Origin and Evolution of Cosmic Structure*. Wiley, 2003.

[57] J.A. Peacock. *Cosmological Physics*. Cambridge Astrophysics. Cambridge University Press, 1999.

[58] E. W. Kolb and M. S. Turner. Book-Review - the Early Universe. S&T, 80:381, October 1990.

[59] T. Padmanabhan. *Structure Formation in the Universe*. Cambridge University Press, 1993.

[60] S. Weinberg. *Cosmology*. Oxford University Press, 2008.

[61] P. J. E. Peebles. *The large-scale structure of the universe*. 1980.

[62] P. J. E. Peebles. *Principles of Physical Cosmology*. 1993.

[63] R. Durrer and R. Maartens. Dark energy and dark gravity: theory overview. *General Relativity and Gravitation*, 40:301–328, February 2008.

[64] S. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23, January 1989.

[65] S. M. Carroll, W. H. Press, and E. L. Turner. The cosmological constant. ARA&A, 30:499–542, 1992.

[66] P. P. Avelino, A. M. M. Trindade, and P. T. P. Viana. Is $w \neq -1$ evidence for a dynamical dark energy equation of state? Phys. Rev. D, 80(6):067302, September 2009.

[67] E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. ApJS, 180:330–376, February 2009.

[68] M. Chevallier and D. Polarski. Accelerating Universes with Scaling Dark Matter. *International Journal of Modern Physics D*, 10:213–223, 2001.

[69] E. V. Linder and A. Jenkins. Cosmic structure growth and dark energy. MNRAS, 346:573–583, December 2003.

[70] J. H. Jeans. The Stability of a Spherical Nebula. *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53, 1902.

[71] E. M. Lifshitz. On the gravitational stability of the expanding universe. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki*, 16:587–602, 1946.

[72] A. J. S. Hamilton. Formulae for growth factors in expanding universes containing matter and a cosmological constant. MNRAS, 322:419–425, April 2001.

[73] E. R. Harrison. Fluctuations at the Threshold of Classical Cosmology. Phys. Rev. D, 1:2726–2730, May 1970.

[74] Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. A&A, 5:84–89, March 1970.

[75] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. Phys. Rep., 367:1–248, September 2002.

[76] T. Padmanabhan. *Structure Formation in the Universe*. May 1993.

[77] S. Cole and C. Lacey. The structure of dark matter haloes in hierarchical clustering models. MNRAS, 281:716, July 1996.

[78] V. R. Eke, S. Cole, and C. S. Frenk. Cluster evolution as a diagnostic for Omega. MNRAS, 282:263–280, September 1996.

[79] V. R. Eke, J. F. Navarro, and C. S. Frenk. The Evolution of X-Ray Clusters in a Low-Density Universe. ApJ, 503:569–592, August 1998.

[80] S. D. M. White, G. Efstathiou, and C. S. Frenk. The amplitude of mass fluctuations in the universe. MNRAS, 262:1023–1028, June 1993.

[81] G. L. Bryan and M. L. Norman. Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. ApJ, 495:80–99, March 1998.

[82] S. F. Shandarin and Y. B. Zeldovich. The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61:185–220, April 1989.

[83] W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ, 187:425–438, February 1974.

[84] K. Jedamzik. The Cloud-in-Cloud Problem in the Press-Schechter Formalism of Hierarchical Structure Formation. ApJ, 448:1, July 1995.

[85] P. P. Avelino and P. T. P. Viana. The cloud-in-cloud problem for non-Gaussian density fields. MNRAS, 314:354–358, May 2000.

[86] R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. MNRAS, 323:1–12, May 2001.

[87] R. W. Hockney and J. W. Eastwood. *Computer Simulation Using Particles*. 1981.

[88] A. W. Appel. An Efficient Program for Many-Body Simulation. *SIAM Journal on Scientific and Statistical Computing, vol. 6, no. 1, January 1985, p. 85-103.*, 6:85–103, January 1985.

[89] J. Barnes and P. Hut. A hierarchical O(N log N) force-calculation algorithm. Nature, 324:446–449, December 1986.

[90] V. Springel, N. Yoshida, and S. D. M. White. GADGET: a code for collisionless and gasdynamical cosmological simulations. New A, 6:79–117, April 2001.

[91] V. Springel. The cosmological simulation code GADGET-2. MNRAS, 364:1105–1134, December 2005.

[92] Viatcheslav Mukhanov. *Physical foundations of cosmology*. Cambridge University Press, 2005.

[93] D. Langlois. Inflation and Cosmological Perturbations. In G. Wolschin, editor, *Lecture Notes in Physics, Berlin Springer Verlag*, volume 800 of *Lecture Notes in Physics, Berlin Springer Verlag*, pages 1–57, March 2010.

[94] A. R. Liddle and D. H. Lyth. The cold dark matter density perturbation. Phys. Rep., 231:1–105, August 1993.

[95] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto. Non-Gaussianity from inflation: theory and observations. Phys. Rep., 402:103–266, November 2004.

[96] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto. Gauge-invariant second-order perturbations and non-Gaussianity from inflation. *Nuclear Physics B*, 667:119–148, September 2003.

[97] E. Komatsu, N. Afshordi, N. Bartolo, D. Baumann, J. R. Bond, E. I. Buchbinder, C. T. Byrnes, X. Chen, D. J. H. Chung, A. Cooray, P. Creminelli, N. Dalal, O. Dore, R. Easther, A. V. Frolov, J. Khoury, W. H. Kinney, L. Kofman, K. Koyama, L. Leblond, J.-L. Lehners, J. E. Lidsey, M. Liguori, E. A. Lim, A. Linde, D. H. Lyth, J. Maldacena, S. Matarrese, L. McAllister, P. McDonald, S. Mukohyama, B. Ovrut, H. V. Peiris, A. Riotto, Y. Rodrigues, M. Sasaki, R. Scoccimarro, D. Seery, A. Sefusatti, K. M. Smith, A. A. Starobinsky, P. J. Steinhardt, F. Takahashi, M. Tegmark, A. J. Tolley, L. Verde, B. D. Wandelt, D. Wands, S. Weinberg, M. Wyman, A. P. S. Yadav, and M. Zaldarriaga. Non-Gaussianity as a Probe of the Physics of the Primordial Universe and the Astrophysics of the Low Redshift Universe. In *astro2010: The Astronomy and Astrophysics Decadal Survey*, ArXiv Astrophysics e-prints, 2009.

[98] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall. Conditions for large non-Gaussianity in two-field slow-roll inflation. J. Cosmology Astropart. Phys., 10:8, October 2008.

[99] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall. Large non-Gaussianity from two-component hybrid inflation. J. Cosmology Astropart. Phys., 2:17, February 2009.

[100] C. T. Byrnes and G. Tasinato. Non-Gaussianity beyond slow roll in multi-field inflation. J. Cosmology Astropart. Phys., 8:16, August 2009.

[101] C. T. Byrnes and K.-Y. Choi. Review of Local Non-Gaussianity from Multifield Inflation. *Advances in Astronomy*, 2010:76, 2010.

[102] N. Bartolo, S. Matarrese, and A. Riotto. Non-Gaussianity in the curvaton scenario. Phys. Rev. D, 69(4):043503, February 2004.

[103] M. Sasaki, J. Väliviita, and D. Wands. Non-Gaussianity of the primordial perturbation in the curvaton model. Phys. Rev. D, 74(10):103003, November 2006.

[104] D. H. Lyth and D. Wands. Generating the curvature perturbation without an inflaton. *Physics Letters B*, 524:5–14, January 2002.

[105] J.-L. Lehners. Ekpyrotic Nongaussianity: A Review. *Advances in Astronomy*, 2010:67, 2010.

[106] J.-L. Lehners and P. J. Steinhardt. Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models. Phys. Rev. D, 77(6):063533, March 2008.

[107] S. Yokoyama and J. Soda. Primordial statistical anisotropy generated at the end of inflation. J. Cosmology Astropart. Phys., 8:5, August 2008.

[108] M. Karčiauskas, K. Dimopoulos, and D. H. Lyth. Anisotropic non-Gaussianity from vector field perturbations. Phys. Rev. D, 80(2):023509, July 2009.

[109] E. Dimastrogiovanni, N. Bartolo, S. Matarrese, and A. Riotto. Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models. *Advances in Astronomy*, 2010:65, 2010.

[110] V. F. Mukhanov and P. J. Steinhardt. Density perturbations in multifield inflationary models. *Physics Letters B*, 422:52–60, March 1998.

[111] D. Polarski and A. A. Starobinsky. Isocurvature perturbations in multiple inflationary models. Phys. Rev. D, 50:6123–6129, November 1994.

[112] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent. Large non-Gaussianity in multiple-field inflation. Phys. Rev. D, 73(8):083522, April 2006.

[113] X. Chen. Primordial Non-Gaussianities from Inflation Models. *Advances in Astronomy*, 2010:72, 2010.

[114] B. Novosyadlyj, R. Durrer, S. Gottlöber, V. N. Lukash, and S. Apunevych. Cosmological parameters from large scale structure observations. A&A, 356:418–434, April 2000.

[115] A. Linde. Hybrid inflation. Phys. Rev. D, 49:748–754, January 1994.

[116] L. Pogosian, S.-H. H. Tye, I. Wasserman, and M. Wyman. Observational constraints on cosmic string production during brane inflation. Phys. Rev. D, 68(2):023506, July 2003.

[117] S. Sarangi and S.-H. H. Tye. Cosmic string production towards the end of brane inflation. *Physics Letters B*, 536:185–192, June 2002.

[118] A. Curto, E. Martínez-González, and R. B. Barreiro. Improved Constraints on Primordial Non-Gaussianity for the Wilkinson Microwave Anisotropy Probe 5-Year Data. ApJ, 706:399–403, November 2009.

[119] A. P. S. Yadav and B. D. Wandelt. Evidence of Primordial Non-Gaussianity (f_{NL}) in the Wilkinson Microwave Anisotropy Probe 3-Year Data at 2.8σ. *Physical Review Letters*, 100(18):181301, May 2008.

[120] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese. An Estimate of the Primordial Non-Gaussianity Parameter f $_{NL}$ Using the Needlet Bispectrum from WMAP. ApJ, 701:369–376, August 2009.

[121] C. Cunha, D. Huterer, and O. Doré. Primordial non-Gaussianity from the covariance of galaxy cluster counts. Phys. Rev. D, 82(2):023004, July 2010.

[122] D. Babich, P. Creminelli, and M. Zaldarriaga. The shape of non-Gaussianities. J. Cosmology Astropart. Phys., 8:9, August 2004.

[123] J. R. Fergusson and E. P. S. Shellard. Shape of primordial non-Gaussianity and the CMB bispectrum. Phys. Rev. D, 80(4):043510, August 2009.

[124] M. Alishahiha, E. Silverstein, and D. Tong. DBI in the sky: Non-Gaussianity from inflation with a speed limit. Phys. Rev. D, 70(12):123505, December 2004.

[125] D. S. Salopek and J. R. Bond. Nonlinear evolution of long-wavelength metric fluctuations in inflationary models. Phys. Rev. D, 42:3936–3962, December 1990.

[126] E. Komatsu and D. N. Spergel. Acoustic signatures in the primary microwave background bispectrum. Phys. Rev. D, 63(6):063002, March 2001.

[127] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach. The three-point correlation function of the cosmic microwave background in inflationary models. ApJ, 430:447–457, August 1994.

[128] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski. Large-scale structure, the cosmic microwave background and primordial non-Gaussianity. MNRAS, 313:141–147, March 2000.

[129] U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. ApJ, 469:437, October 1996.

[130] M. Zaldarriaga, U. Seljak, and E. Bertschinger. Integral Solution for the Microwave Background Anisotropies in Nonflat Universes. ApJ, 494:491–502, February 1998.

[131] M. Zaldarriaga and U. Seljak. CMBFAST for Spatially Closed Universes. ApJS, 129:431–434, August 2000.

[132] A. Lewis, A. Challinor, and A. Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. ApJ, 538:473–476, August 2000.

[133] M. Doran. CMBEASY: an object oriented code for the cosmic microwave background. J. Cosmology Astropart. Phys., 10:11, October 2005.

[134] J. Lesgourgues. The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *ArXiv e-prints*, April 2011.

[135] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. ApJ, 304:15–61, May 1986.

[136] N. Sugiyama. Cosmic Background Anisotropies in Cold Dark Matter Cosmology. ApJS, 100:281, October 1995.

[137] S. Dodelson, G. Gyuk, and M. S. Turner. Is a massive tau neutrino just what cold dark matter needs\? *Physical Review Letters*, 72:3754–3757, June 1994.

[138] R. Juszkiewicz, D. H. Weinberg, P. Amsterdamski, M. Chodorowski, and F. Bouchet. Weakly nonlinear Gaussian fluctuations and the edgeworth expansion. ApJ, 442:39–56, March 1995.

[139] F. Bernardeau and L. Kofman. Properties of the cosmological density distribution function. ApJ, 443:479–498, April 1995.

[140] M. Abramowitz, I. Stegun, and N. Sfetcu. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Nicolae Sfetcu, 2014.

[141] A. M. M. Trindade, P. P. Avelino, and P. T. P. Viana. A new signature of primordial non-Gaussianities from the abundance of galaxy clusters. MNRAS, 424:1442–1447, August 2012.

[142] T. Giannantonio and C. Porciani. Structure formation from non-Gaussian initial conditions: Multivariate biasing, statistics, and comparison with N-body simulations. Phys. Rev. D, 81(6):063530, March 2010.

[143] M. Maggiore and A. Riotto. The Halo Mass Function from Excursion Set Theory. III. Non-Gaussian Fluctuations. ApJ, 717:526–541, July 2010.

[144] G. D'Amico, M. Musso, J. Noreña, and A. Paranjape. An improved calculation of the non-Gaussian halo mass function. J. Cosmology Astropart. Phys., 2:001, February 2011.

[145] M. LoVerde and K. M. Smith. The non-Gaussian halo mass function with f_{NL}, g_{NL} and τ_{NL}. J. Cosmology Astropart. Phys., 8:003, August 2011.

[146] I. E. Achitouv and P. S. Corasaniti. Non-Gaussian halo mass function and non-spherical halo collapse: theory vs. simulations. J. Cosmology Astropart. Phys., 2:002, February 2012.

[147] A. D'Aloisio, J. Zhang, D. Jeong, and P. R. Shapiro. Halo statistics in non-Gaussian cosmologies: the collapsed fraction, conditional mass function and halo bias from the path-integral excursion set method. MNRAS, 428:2765–2788, January 2013.

[148] P. Valageas, N. Clerc, F. Pacaud, and M. Pierre. Covariance matrices for halo number counts and correlation functions. A&A, 536:A95, December 2011.

[149] M. Sahlen, P. T. P. Viana, A. R. Liddle, A. K. Romer, M. Davidson, M. Hosmer, E. Lloyd-Davies, K. Sabirli, C. A. Collins, P. E. Freeman, M. Hilton, B. Hoyle, S. T. Kay, R. G. Mann, N. Mehrtens, C. J. Miller, R. C. Nichol, S. A. Stanford, and M. J. West. The xmm cluster survey: forecasting cosmological and cluster scaling-relation parameter constraints. *Mon. Not. R. Astron. Soc.*, 397:577–607, 2009.

[150] U. von Toussaint. Bayesian inference in physics. *Reviews of Modern Physics*, 83:943–999, July 2011.

[151] S. Shandera, A. Mantz, D. Rapetti, and S. W. Allen. X-ray cluster constraints on non-Gaussianity. J. Cosmology Astropart. Phys., 8:004, August 2013.

[152] A. M. M. Trindade, P. P. Avelino, and P. T. P. Viana. Biased cosmological parameter estimation with galaxy cluster counts in the presence of primordial non-Gaussianities. MNRAS, 435:782–785, October 2013.

[153] S. Lim and J. Lee. The Extended Zel'dovich Mass Functions of Clusters and Isolated Clusters in the Presence of Primordial Non-Gaussianity. ApJ, 792:91, September 2014.

[154] X. Kang, P. Norberg, and J. Silk. Can a large-scale structure probe cosmic microwave background-constrained non-Gaussianity? MNRAS, 376:343–347, March 2007.

[155] M. Grossi, K. Dolag, E. Branchini, S. Matarrese, and L. Moscardini. Evolution of massive haloes in non-Gaussian scenarios. MNRAS, 382:1261–1267, December 2007.

[156] M. Grossi, L. Verde, C. Carbone, K. Dolag, E. Branchini, F. Iannuzzi, S. Matarrese, and L. Moscardini. Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations. MNRAS, 398:321–332, September 2009.

[157] V. Desjacques, U. Seljak, and I. T. Iliev. Scale-dependent bias induced by local non-Gaussianity: a comparison to N-body simulations. MNRAS, 396:85–96, June 2009.

[158] A. Pillepich, C. Porciani, and O. Hahn. Halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions. MNRAS, 402:191–206, February 2010.

[159] C. Wagner, L. Verde, and L. Boubekeur. N-body simulations with generic non-Gaussian initial conditions I: power spectrum and halo mass function. J. Cosmology Astropart. Phys., 10:022, October 2010.

[160] R. E. Smith, V. Desjacques, and L. Marian. Nonlinear clustering in models with primordial non-Gaussianity: The halo model approach. Phys. Rev. D, 83(4):043526, February 2011.

[161] C. Wagner and L. Verde. N-body simulations with generic non-Gaussian initial conditions II: halo bias. J. Cosmology Astropart. Phys., 3:002, March 2012.

[162] U. Maio and F. Iannuzzi. Baryon history and cosmic star formation in non-Gaussian cosmological models: numerical simulations. MNRAS, 415:3021–3032, August 2011.

[163] X. Zhao, Y. Li, S. Shandera, and D. Jeong. The Effects of Local Primordial Non-Gaussianity on the Formation and Evolution of Galaxies. *ArXiv e-prints*, July 2013.

[164] F. Pace and U. Maio. Hydrodynamical chemistry simulations of the Sunyaev-Zel'dovich effect and the impacts from primordial non-Gaussianities. MNRAS, 437:1308–1317, January 2014.

[165] B. Sartoris, S. Borgani, C. Fedeli, S. Matarrese, L. Moscardini, P. Rosati, and J. Weller. The potential of X-ray cluster surveys to constrain primordial non-Gaussianity. MNRAS, 407:2339–2354, October 2010.

[166] M. Roncarelli, L. Moscardini, E. Branchini, K. Dolag, M. Grossi, F. Iannuzzi, and S. Matarrese. Imprints of primordial non-Gaussianities in X-ray and SZ signals from galaxy clusters. MNRAS, 402:923–933, February 2010.

[167] A. Pillepich, C. Porciani, and T. H. Reiprich. The X-ray cluster survey with eRosita: forecasts for cosmology, cluster physics and primordial non-Gaussianity. MNRAS, 422:44–69, May 2012.

[168] D. S. Y. Mak and E. Pierpaoli. Constraints on non-Gaussianity from Sunyaev-Zeldovich cluster surveys. Phys. Rev. D, 86(12):123520, December 2012.

[169] S. Khedekar and S. Majumdar. Cosmology with the largest galaxy cluster surveys: going beyond Fisher matrix forecasts. J. Cosmology Astropart. Phys., 2:030, February 2013.

[170] N. Kaiser. Evolution and clustering of rich clusters. MNRAS, 222:323–345, September 1986.

[171] A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. ARA&A, 50:353–409, September 2012.

[172] A. Moradinezhad Dizgah, S. Dodelson, and A. Riotto. Imprint of primordial non-Gaussianity on dark matter halo profiles. Phys. Rev. D, 88(6):063513, September 2013.

[173] R. Scoccimarro, L. Hui, M. Manera, and K. C. Chan. Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity. Phys. Rev. D, 85(8):083002, April 2012.

[174] A. Lewis. CAMB Notes 2014. 2014.

[175] P. A. Thomas, J. M. Colberg, H. M. P. Couchman, G. P. Efstathiou, C. S. Frenk, A. R. Jenkins, A. H. Nelson, R. M. Hutchings, J. A. Peacock, F. R. Pearce, S. D. M. White, and Virgo Consortium. The structure of galaxy clusters in various cosmologies. MNRAS, 296:1061–1071, June 1998.

[176] F. R. Pearce, P. A. Thomas, H. M. P. Couchman, and A. C. Edge. The effect of radiative cooling on the X-ray properties of galaxy clusters. MNRAS, 317:1029–1040, October 2000.

[177] O. Muanwong, P. A. Thomas, S. T. Kay, F. R. Pearce, and H. M. P. Couchman. The Effect of Radiative Cooling on Scaling Laws of X-Ray Groups and Clusters. ApJ, 552:L27–L30, May 2001.

[178] A. C. da Silva, S. T. Kay, A. R. Liddle, and P. A. Thomas. Hydrodynamical simulations of the Sunyaev-Zel'dovich effect: cluster scaling relations and X-ray properties. MNRAS, 348:1401–1408, March 2004.

[179] R. S. Sutherland and M. A. Dopita. Cooling functions for low-density astrophysical plasmas. ApJS, 88:253–327, September 1993.

[180] A. C. da Silva, A. Catalano, L. Montier, E. Pointecouteau, J. Lanoux, and M. Giard. The impact of dust on the scaling properties of galaxy clusters. MNRAS, 396:849–859, June 2009.

[181] N. Aghanim, A. C. da Silva, and N. J. Nunes. Cluster scaling relations from cosmological hydrodynamic simulations in a dark-energy dominated universe. A&A, 496:637–644, March 2009.

[182] A. Riotto and M. S. Sloth. Strongly scale-dependent non-Gaussianity. Phys. Rev. D, 83(4):041301, February 2011.

[183] Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, C. Baccigalupi, A. J. Banday, and et al. Planck 2013 results. XX. Cosmology from Sunyaev-Zeldovich cluster counts. A&A, 571:A20, November 2014.

[184] S. T. Kay, A. C. da Silva, N. Aghanim, A. Blanchard, A. R. Liddle, J.-L. Puget, R. Sadat, and P. A. Thomas. The evolution of clusters in the CLEF cosmological simulation: X-ray structural and scaling properties. MNRAS, 377:317–334, May 2007.

[185] K. Borne. Scientific Data Mining in Astronomy. *Taylor & Francis: CRC Press, Boca Raton, FL, Ch.*, 5:99–114, November 2009.

[186] N. M. Ball and R. J. Brunner. Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics D*, 19:1049–1106, 2010.

[187] R. S. de Souza, U. Maio, V. Biffi, and B. Ciardi. Robust PCA and MIC statistics of baryons in early minihaloes. MNRAS, 440:240–248, May 2014.

[188] R. S. de Souza and B. Ciardi. AMADA-Analysis of multidimensional astronomical datasets. *Astronomy and Computing*, 12:100–108, September 2015.

[189] John K Kruschke. *Doing Bayesian data analysis : a tutorial with R and BUGS*. Academic Press, Burlington, MA, 2011.

[190] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. J. Chem. Phys., 21:1087–1092, June 1953.

[191] H. Sana, S. E. de Mink, A. de Koter, N. Langer, C. J. Evans, M. Gieles, E. Gosset, R. G. Izzard, J.-B. Le Bouquin, and F. R. N. Schneider. Binary Interaction Dominates the Evolution of Massive Stars. *Science*, 337:444–, July 2012.

[192] M. Janson, C. Bergfors, W. Brandner, N. Kudryavtseva, F. Hormuth, S. Hippler, and T. Henning. The AstraLux Multiplicity Survey: Extension to Late M-dwarfs. ApJ, 789:102, July 2014.

[193] P. Uttley, I. M. McHardy, and I. E. Papadakis. Measuring the broad-band power spectra of active galactic nuclei with RXTE. MNRAS, 332:231–250, May 2002.

[194] T. T. Shimizu and R. F. Mushotzky. The First Hard X-Ray Power Spectral Density Functions of Active Galactic Nucleus. ApJ, 770:60, June 2013.

[195] V. L. Kashyap, J. J. Drake, M. Güdel, and M. Audard. Flare Heating in Stellar Coronae. ApJ, 580:1118–1132, December 2002.

[196] Donald B. Rubin. Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, 12(4):pp. 1151–1172, 1984.

[197] S. Tavaré, D. Balding, R. Griffith, and P. Donnelly. Inferring coalescence times from DNA sequence data. *Genetics*, 145:505, 1997.

[198] Jonathan K Pritchard, Mark T Seielstad, Anna Perez-Lezaun, and Marcus W Feldman. Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999.

[199] E. Cameron and A. N. Pettitt. Approximate Bayesian Computation for astronomical model analysis: a case study in galaxy demographics and morphological transformation at high redshift. MNRAS, 425:44–65, September 2012.

[200] ChadM. Schafer and PeterE. Freeman. Likelihood-free inference in cosmology: Potential for the estimation of luminosity functions. In Eric D. Feigelson and G. Jogesh Babu, editors, *Statistical Challenges in Modern Astronomy V*, Lecture Notes in Statistics, pages 3–19. Springer New York, 2012.

[201] A. Weyant, C. Schafer, and W. M. Wood-Vasey. Likelihood-free Cosmological Inference with Type Ia Supernovae: Approximate Bayesian Computation for a Complete Treatment of Uncertainty. ApJ, 764:116, February 2013.

[202] A. C. Robin, C. Reyle, J. Fliri, M. Czekaj, C. P. Robert, and A. M. M. Martins. Constraining the thick disc formation scenario of the Milky Way. *ArXiv e-prints*, June 2014.

[203] Paul Marjoram, John Molitor, Vincent Plagnol, and Simon Tavaré. Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.

[204] Mark A Beaumont, Jean-Marie Cornuet, Jean-Michel Marin, and Christian P Robert. Adaptive approximate bayesian computation. *Biometrika*, page asp052, 2009.

[205] C.-A. Lin and M. Kilbinger. A new model to predict weak-lensing peak counts II. Parameter constraint strategies. *ArXiv e-prints*, June 2015.

[206] M. Killedar, S. Borgani, D. Fabjan, K. Dolag, G. L. Granato, M. Meneghetti, S. Planelles, and C. Ragone-Figueroa. Weighted ABC: a new strategy for cluster strong lensing cosmology with simulations. *ArXiv e-prints*, July 2015.

[207] J. Akeret, A. Refregier, A. Amara, S. Seehars, and C. Hasner. Approximate Bayesian computation for forward modeling in cosmology. *JCAP*, 8:43, August 2015.

[208] Juliane Liepe, Chris Barnes, Erika Cule, Kamil Erguler, Paul Kirk, Tina Toni, and Michael P.H. Stumpf. Abc-sysbio—approximate bayesian computation in python with gpu support. *Bioinformatics*, 26(14):1797–1799, 2010.

[209] J. R. Oaks. An Improved Approximate-Bayesian Model-choice Method for Estimating Shared Evolutionary History. *ArXiv e-prints*, February 2014.

[210] S. D. P. Vitenti and M. Penna-Lima. NumCosmo: Numerical Cosmology. *ASCL:1408.013*, August 2014.

[211] S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1760–1765, 2007.

[212] C. C. Drovandi and A. N. Pettitt. Estimation of parameters for macroparasite population evolution using approximate bayesian computation. *Biometrics*, 67(1):225–233, 2011.

[213] Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[214] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. An adaptive sequential monte carlo method for approximate bayesian computation. *Statistics and Computing*, 22(5):1009–1020, 2012.

[215] Oliver Ratmann, Anton Camacho, Adam Meijer, and Gé Donker. Statistical modelling of summary values leads to accurate approximate bayesian computations. *arXiv preprint arXiv:1305.4283*, 2013.

[216] Mark A Beaumont, Jean-Marie Cornuet, Jean-Michel Marin, and Christian P Robert. Adaptive approximate bayesian computation. *Biometrika*, page asp052, 2009.

[217] Paul Fearnhead and Dennis Prangle. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.

[218] M. G. B. Blum, M. A. Nunes, D. Prangle, and S. A. Sisson. A comparative review of dimension reduction methods in approximate bayesian computation. *Statist. Sci.*, 28(2):189–208, 05 2013.

[219] M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, and et al. Cosmological constraints from the SDSS luminous red galaxies. Phys. Rev. D, 74(12):123507, December 2006.

[220] A. J. Benson. Galaxy formation theory. Phys. Rep., 495:33–86, October 2010.

[221] A. V. Kravtsov and S. Borgani. Formation of Galaxy Clusters. ARA&A, 50:353–409, September 2012.

[222] S. W. Allen, D. A. Rapetti, R. W. Schmidt, H. Ebeling, R. G. Morris, and A. C. Fabian. Improved constraints on dark energy from Chandra X-ray observations of the largest relaxed galaxy clusters. MNRAS, 383:879–896, January 2008.

[223] M. Penna-Lima, M. Makler, and C. A. Wuensche. Biases on cosmological parameter estimators from galaxy cluster number counts. *JCAP*, 5:39, May 2014.

[224] T. F. Laganá, R. S. de Souza, and G. R. Keller. On the influence of non-thermal pressure on the mass determination of galaxy clusters. A&A, 510:A76, February 2010.

[225] S. Giodini, L. Lovisari, E. Pointecouteau, S. Ettori, T. H. Reiprich, and H. Hoekstra. Scaling relations for galaxy clusters: Properties and evolution. *Space Science Reviews*, June 2013.

[226] M. Birkinshaw. The Sunyaev-Zel'dovich effect. Phys. Rep., 310:97–195, March 1999.

[227] J. E. Carlstrom, G. P. Holder, and E. D. Reese. Cosmology with the Sunyaev-Zel'dovich Effect. ARA&A, 40:643–680, 2002.

[228] D. Barbosa, J. G. Bartlett, A. Blanchard, and J. Oukbir. The Sunyaev-Zel'dovich effect and the value of {OMEGA}_0_. A&A, 314:13–17, October 1996.

[229] P. M. Motl, E. J. Hallman, J. O. Burns, and M. L. Norman. The Integrated Sunyaev-Zeldovich Effect as a Superior Method for Measuring the Mass of Clusters of Galaxies. ApJ, 623:L63–L66, April 2005.

[230] K. Vanderlinde, T. M. Crawford, T. de Haan, J. P. Dudley, L. Shaw, P. A. R. Ade, K. A. Aird, B. A. Benson, and et al. Galaxy Clusters Selected with the Sunyaev-Zel'dovich Effect from 2008 South Pole Telescope Observations. ApJ, 722:1180–1196, October 2010.

[231] B. A. Benson, T. de Haan, J. P. Dudley, C. L. Reichardt, K. A. Aird, K. Andersson, R. Armstrong, M. L. N. Ashby, and et al. Cosmological Constraints from Sunyaev-Zel'dovich-selected Clusters with X-Ray Observations in the First 178 deg^2 of the South Pole Telescope Survey. ApJ, 763:147, February 2013.

[232] C. L. Reichardt, B. Stalder, L. E. Bleem, T. E. Montroy, K. A. Aird, K. Andersson, R. Armstrong, M. L. N. Ashby, and et al. Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the First 720 Square Degrees of the South Pole Telescope Survey. ApJ, 763:127, February 2013.

[233] L. E. Bleem, B. Stalder, T. de Haan, K. A. Aird, S. W. Allen, D. E. Applegate, M. L. N. Ashby, Bautz, and et. al. Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey. ApJS, 216:27, February 2015.

[234] Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 1st edition, 2008.

[235] C. Fedeli, C. Carbone, L. Moscardini, and A. Cimatti. The clustering of galaxies and galaxy clusters: constraints on primordial non-Gaussianity from future wide-field surveys. MNRAS, 414:1545–1559, June 2011.

[236] T. Giannantonio, A. J. Ross, W. J. Percival, R. Crittenden, D. Bacher, M. Kilbinger, R. Nichol, and J. Weller. Improved primordial non-Gaussianity constraints from measurements of galaxy clustering and the integrated Sachs-Wolfe effect. Phys. Rev. D, 89(2):023511, January 2014.

[237] S. Gariazzo, L. Lopez-Honorez, and O. Mena. Primordial power spectrum features and f_{NL} constraints. Phys. Rev. D, 92(6):063510, September 2015.

[238] B. Kalus, W. J. Percival, and L. Samushia. Cosmological parameter inference from galaxy clustering: the effect of the posterior distribution of the power spectrum. MNRAS, 455:2573–2581, January 2016.

[239] C. Fedeli, M. Bartelmann, and L. Moscardini. Constraining primordial non-Gaussianity with cosmological weak lensing: shear and flexion. J. Cosmology Astropart. Phys., 10:018, October 2012.

[240] L. Marian, S. Hilbert, R. E. Smith, P. Schneider, and V. Desjacques. Measuring Primordial Non-gaussianity Through Weak-lensing Peak Counts. ApJ, 728:L13, February 2011.

[241] M. Maturi, C. Fedeli, and L. Moscardini. Imprints of primordial non-Gaussianity on the number counts of cosmic shear peaks. MNRAS, 416:2527–2538, October 2011.

[242] C. Fedeli and L. Moscardini. Cosmic shear statistics in cosmologies with non-Gaussian initial conditions. MNRAS, 405:681–694, June 2010.

[243] A. Cooray. 21-cm Background Anisotropies Can Discern Primordial Non-Gaussianity. *Physical Review Letters*, 97(26):261301, December 2006.

[244] S. Joudaki, O. Doré, L. Ferramacho, M. Kaplinghat, and M. G. Santos. Primordial Non-Gaussianity from the 21 cm Power Spectrum during the Epoch of Reionization. *Physical Review Letters*, 107(13):131304, September 2011.

[245] S. Chongchitnan and J. Silk. The 21-cm radiation from minihaloes as a probe of small primordial non-Gaussianity. MNRAS, 426:L21–L25, October 2012.